

**Opening a Can of Worms: Idiom Flexibility, Decomposability, and
the Mental Lexicon**

By

Zachary Maher

Advisor: Maria Mercedes Piñango

Submitted to the faculty of the Department of Linguistics
in partial fulfillment of the requirements for the degrees of
Bachelor of Arts and Master of Arts

Yale University

1 May 2013

In A Nutshell

Despite their prevalence in the English language, idiomatic expressions (e.g. kick the bucket, tip of the iceberg) have received only limited treatment in the linguistic literature. In this thesis, I ask the question of what speakers need to know in order to be competent users and comprehenders of idioms. I focus particularly on idioms' variant flexibility (e.g. *Some strings were pulled by the politician* vs. *#The bucket was kicked by the man*). Many different frameworks have been proposed over small sets of idealized data, resulting in conclusions that do not truly capture the behavior of idioms. I present the first study to look at a group of syntactically diverse idioms, gathering acceptability judgments for each idiom presented separately with 3-4 modifications (adjective insertion, two types of word substitution, passivization). Results indicate that previous knowledge of an idiom does not affect the reliability with which participants deem a modification acceptable, suggesting that idiom syntactic flexibility is not encoded idiosyncratically in the lexicon. Additionally, idiom flexibility is not a single spectrum; an idiom can accept adjective insertion but not word substitution or vice versa, as observed through regression analysis of the judgment data. Finally, the presence of a large set of judgment data allows us to begin to propose new principles to explain idiom flexibility. A follow-up study addresses a confound in Gibbs & Nayak's (1989) Idiom Decomposability Hypothesis, one of the most influential early models of idiom flexibility, and suggests that participants categorize idioms according to their own internalized definitions, rather than being influenced by what is presented to them. Altogether, I present the case that idioms are represented in the mental lexicon, but most of their flexibility is explained by principles operating at the the syntax/semantics and semantics/conceptual structure interfaces.

Contents

1	Starting From Square One	6
1.1	Introduction and Roadmap	6
1.2	Definitions	8
1.3	Idiom Flexibility	9
1.4	Accounts	11
1.4.1	Decomposability	11
1.4.2	Configuration Hypothesis	14
1.4.3	Comparing the Models	15
1.5	Open Issues	18
2	Experiment 1: Opening the Can of Worms	19
2.1	Background and Purpose	19
2.2	Scalar Substitution	20
2.3	Design and Methodology	21
2.3.1	General Design	21
2.3.2	Amazon Mechanical Turk	21
2.3.3	Participants	22
2.3.4	Materials	22
2.3.5	Procedure	25
2.4	Results	26
2.4.1	Quality Control	27
2.4.2	Judgments	27
2.4.3	Judgment Consistency	31
2.4.4	Judgment Patterns	33
2.4.5	Examination of Items	38
2.5	Discussion	39
2.6	Conclusion	41
3	Experiment 2: Skeletons in the Closet	43
3.1	Background	43

3.2	Experiment 2A	45
3.2.1	Methods	46
3.2.2	Results	48
3.2.3	Discussion	51
3.3	Experiment 2B	52
3.3.1	Method	52
3.3.2	Procedure	53
3.3.3	Results	54
3.3.4	Discussion	56
3.4	Pooled Analysis	57
3.5	General Discussion	57
3.6	Conclusions	59
4	Where the Rubber Meets the Road	60
4.1	Idioms and Collocations	60
4.1.1	Collocations containing Idioms	60
4.1.2	Light Verb Idioms	61
4.1.3	Implications for Flexibility	63
4.2	Conceptual Metaphors, Conceptual Structure	66
4.2.1	Idiom Analyzability	66
4.3	The Lexicon and Flexibility	68
4.4	Conclusion	70
5	Wrapping Up	72
A	Experiment 1 Stimuli	75
A.1	Real	75
A.1.1	VP	75
A.1.2	PP	79
A.1.3	NP	82
A.2	Novel Idioms	84
A.2.1	VP	84
A.2.2	PP	89
A.2.3	NP	91
B	Results of Experiment 1	95
C	Plausibility for Novel Idioms in Experiment 1	98
C.1	Overview	98
C.1.1	Participants	98
C.1.2	Materials and Procedure	98

- C.2 Results 99
 - C.2.1 Quality Control 99
 - C.2.2 Plausibility Scores 99

Chapter 1

Starting From Square One

1.1 Introduction and Roadmap

This thesis proceeds from the following broad question: what does a speaker need to know in order to properly use and understand an idiom? Many approaches could be taken in answering this question, but this thesis will focus on two specific questions: what about idioms is and is not stored in the lexicon, and what is the explanatory power of Gibbs and Nayak’s (1989) Idiom Decomposability Hypothesis? Here, “properly use” can be considered shorthand for a speaker’s competency with idiom flexibility. Thus, we could think of this as asking what idiosyncratic information and principled information speakers need to know in order to accept passivization for *pull strings* but not for *kick the bucket* (1).

- (1) a. Some **strings were pulled** to get him the job.
- b. #**The bucket was kicked** by the old man last night, so it’s time to plan a funeral.

Answering the broad question through the specific questions is crucial to a well-defined model of the human language faculty, since idioms pervade language. By Jackendoff’s estimation, they are as numerous as adjectives in English. However, in the linguistic literature, they have received only limited attention. Often, it seems that we default to the concept that idioms are lexical items just like *dog*, *diaper*, and *delineate*. Such an approach can serve as a reasonable null hypothesis, and when unqualified, this approach would lead us to the following two conclusions: (1) idioms’ representation in the mental lexicon is essentially the same as that of any vanilla-flavored word, and (2) idioms are inflexible, meaning they cannot undergo syntactic transformations or accept the word insertions and substitutions.

As we shall soon see, at least the latter conclusion contradicts countless attestations of modified idioms. Many accounts have been proposed to explain certain patterns of idiom flexibility, but

they face the same issue that the proponents of the null hypothesis faced: misgeneralization from idealized cases, assuming that all idioms behave as do a particular set under examination.

In order to ask what a speaker needs to know in order to properly use an idiom, this thesis examines the question of what it means to reject the null hypothesis. As my title suggestion, this endeavor unavoidably opens a can of worms. Whereas others have simply opened this can, taken a look at what's on top, and given the best description, I seek to dig through the contents of the can and understand a representative set of worms. In this research program, I face the risk of plodding through data endlessly without coming to any coherent framework, an approach that would perhaps be the most honest to the complex data out there, but which offers little in terms of steps forward in idiom research. Instead, I seek a way forward in which we try to understand exactly what it means to reject the null hypothesis. I do this by asking two questions. First, what factors and patterns explain the existence of idiom flexibility (particularly, how much flexibility can we derive from decomposability)? Second, what do these patterns tell us about the lexical status or non-status of idioms?

The present chapter provides the necessary background to begin the endeavor. Most importantly, we need to develop a definition for idioms before we can study them. After this, I will outline the evidence that some idioms are flexible for certain operations, leading us to reject the null hypothesis that idioms are chunks with no internal analysis. I will conclude the introduction with two proposed models of idiomaticity, one focusing on a principle for flexibility (the Idiom Decomposability Hypothesis) and the other focusing on idioms' representation (the Configuration Hypothesis). By examining the claims of these hypotheses, we will begin to determine what speaker competence in analyzing an idiom requires and what exactly needs to be stored for a speaker to "know an idiom."

Since idioms are flexible, we need to know where this flexibility comes from. Does the rejection of the second conclusion of the null hypothesis require that we also reject the first conclusion? On the contrary, some proposals suggest that flexibility is specified within an idiom's lexical item. Experiment 1 (Chapter 2) confirms a study that refutes this claim, correcting for the non-representativeness problem. On the other hand, are there clear principles underlying idiom flexibility? Experiment 1 uses an expanded data set to evaluate some existing proposals, suggesting that idiom flexibility is far too complex to be explained by a handful of simple principles. Particularly, decomposability as it has been classically defined provides fairly limited explanatory power. However, the more we understand those principles, the more we can understand how idiom flexibility interacts with the lexicon, so evaluation of existing models and proposals for new models becomes germane to understanding what we need to know in order to use an idiom properly.

Experiment 2 (Chapter 3) addresses a confound in a seminal study (Gibbs and Nayak, 1989), a confound which could undermine the existence of their proposed explanatory factor of decomposability. In addressing this confound, I ask whether speakers can be forced to analyze an idiom differently, a question which bears on the rigidity of the representations in the lexicon – a crucial

part of what speakers need to know in order to properly understand an idiom. The results of both experiments will be discussed together in Chapter 4, focusing on the question of what the data tell us about the lexical status of idioms and principles underlying their flexibility. Chapter 5 will provide a summary, conclusions, and proposals for future work.

1.2 Definitions

Before we can begin to discuss the details of idiomaticity, we must first develop a working definition of the term *idiom*. Colloquially, this term can cover a broad range of expressions, including clichés and collocations such as those in (2). Anecdotally, I have found that most non-linguists, when confronted with the term “idiom,” think of adages (3). In short, we might think of all of the multi-word expressions that tend to be solutions in the popular game show *Wheel of Fortune* (Jackendoff, 1995).

- (2) a. trick or treat (cf. ?#treat or trick).
- b. no money down
- c. frequent flyer program
- (3) a. A penny saved is a penny earned.
- b. The early bird gets the worm.
- c. You can’t have your cake and eat it, too.

The linguistic literature on idiomaticity usually adopts a more specific, albeit still imprecise, definition of the term *idiom*, exemplified well by the *OED*: “a group of words established by usage as having a meaning not deducible from the meanings of the individual words” (sense 3). Thus, idioms are a particular type of fixed expression that distinguish themselves from collocations and clichés in that the basic composition of their individual words would not allow an otherwise competent speaker of a language to understand the meaning (barring other contextual clues, of course).

Nunberg et al. (1994) provides a more thorough definition describing the properties of a prototypical idiom, rather than seeking a clean definition. Such an approach makes sense given the presence of edge cases, including conventionalized metaphors such as *bridge the gap*. Novel metaphors appear all the time in everyday speech, and we understand them through metaphorical understanding of the literal meaning of the phrase. Thus, might expect someone who had never before heard *bridge the gap* to understand (4a), and we might also expect anyone to derive a similar metaphorical meaning from the non-conventionalized metaphor *close the rift* in (4b).

- (4) a. Dr. Jones’s research **bridges the gap** between formal syntax and psycholinguistics.
- b. Dr. Jones’s research **closes the rift** between formal syntax and psycholinguistics.

My personal inclination is still to categorize *bridge the gap* as an idiom, but it exemplifies the blurriness of the category. Beyond this, the adages in (3) similarly might be understood metaphorically without prior knowledge. Indeed, we can translate foreign adages, understand them, and ponder them. Since I will not be examining sentential idioms, it is not presently necessary for clean categorization of the examples in (3), but the line between frozen metaphors and adages is itself blurry and perhaps artificial.

Once we acknowledge that idiomaticity truly represents a murky range of a multidimensional spectrum, we might be inclined to “throw our hands in the air” and deem the term meaningless. However, we must remember that there are many expressions that all researchers wish to label as idioms, including the infamous *kick the bucket*, as well as others such as *can of worms*, *out of the blue*, and *pull strings*. Indeed, enough of these terms exist to have already facilitated decades of research. More importantly, they remain an important starting point in pursuing work on formulaic language. Because there exist some phrases that we *must* know wholesale, we can immediately assume that they must be stored and represented in something that at least resembles the mental lexicon. Thus, a focus on idioms provides some bedrock for our understanding of phrasal language. The fact that the target’s edges are ill-defined does preclude us from studying the bullseye of prototypical idioms.

Thus, for the remainder of this paper, I will use the term *idiom* to refer to non-compositional phrases that must be remembered wholesale. I use this term with free admission of its theoretical imprecision, and I hope to continue to “push the boundaries” through my work.

1.3 Idiom Flexibility

The idea that at least the most prototypical of idioms must be lexical might lead us to conclude that idioms are simply a marginal part of the lexicon, capable of receiving no more internal modification than words. This was particularly the view of Chomsky (1981), who suggested that idioms like *kick the bucket* are lexical verbs, inserting at a V^0 head; his notion of the lexicon does not permit phrases, leading to the representation replicated in (5):



Even some psycholinguistic models have treated idioms as morphologically complex words, e.g. the Lexical Representation Hypothesis (Swinney, 1979). Such an analysis might be consistent with the data for *kick the bucket*, but it does not capture the considerable amount of flexibility that many other idioms exhibit. For example, in the naturally-occurring examples below, we can see insertion of both adjectives (6a) and nouns (6b). We also observe different types of word substitutions, such as a change of determiner from *call the shots* to *call this shot* (7a), and a more complex substitution in (7b). In this latter substitution, adapted from *V one's way to the top*, the word *middle* replaces *top*, generating a new and distinct meaning from the original idiom.

- (6) a. “more bang for your **educational** buck” (Obama, State of the Union 2013)
 b. “all the times Michelle Williams got the short end of the **Destiny's Child** stick”
 (poormichelle.com)
- (7) a. “That’s why I’m calling **this** shot.” (Columbo, S3E1)
 b. “It’s not worth [sleep]ing your way to the **middle**.” (House of Cards, E9)

For verb phrase idioms, a whole host of syntactic manipulations are possible, including passivization (8a), clefting (8b), quantification (8c), topicalization (8d), and ellipsis (8e), which have been thoroughly documented by Nunberg et al. (1994). Curiously, however, other idioms with the same syntax ([_{VP} V NP]) are inflexible for the same manipulations (9).¹

- (8) Flexible VP Idioms
- a. The journalist’s **thunder was stolen** by the New York Times when she got scooped.
 (Passivization)
- b. The **strings that he pulled** helped him to gain admission to the prestigious school.
 (Clefting)
- c. I pulled **some strings** to get the job. (Quantification)
- d. **Those strings**, he wouldn’t **pull** for you. (Nunberg et al. (1994); Topicalization)
- e. My goose is cooked, but **yours** isn’t. (Nunberg et al. (1994); Ellipsis)
- (9) Inflexible VP Idioms
- a. # **The bucket was kicked** by the old man last night. (Passivization)
- b. #**The beans that Jack spilled** were far more confidential than he realized. (Clefting)
- c. #Rick **cleared some of the air**, but not all of it. (Quantification)
- d. #**The beans**, she will never **spill**. (Topicalization)
- e. #The college students **hit the sauce** on weekends, and they also **hit it** during makeup classes. (Ellipsis)

¹Throughout this paper, I will use the symbol “#,” which generally refers to semantically or pragmatically infelicitous, though grammatical, sentences to indicate that the idiomatic reading is unavailable. For example, (9a) is a perfectly acceptable sentence if a literal bucket is being kicked, but the idiomatic meaning ‘to die’ is unavailable.

Despite Chomsky’s marginalization of the phenomenon of idiomaticity, these data suggest that idioms play an important role in the productive linguistic system. However, they challenge our traditional notions of compositionality, since they do not seem to be as compositional as their literal phrasal counterparts, but they are also more compositional than opaque chunks. At least in the English language, formulaic language is very common. The continued existence of *Wheel of Fortune* testifies to this. Therefore, if the purpose of the linguistic research program is indeed to understand the human language faculty, linguists should prioritize the pursuit of understanding of these puzzling patterns. These puzzling patterns form the core of what I seek to explain in this thesis. My question of “proper use” of idioms boils down to a question of the source of speakers’ judgments regarding the flexibility of idioms in different constructions.

1.4 Accounts

Over the past four decades, many attempts have been made to explain the varying degrees of idiom flexibility. Fraser (1970) provides the earliest theoretical account to my knowledge, where he proposed different categories of idioms corresponding to different levels of flexibility within the framework of Transformational Grammar. However, the most persistent ideas in the literature have been decomposability (Nunberg, 1978; Gibbs and Nayak, 1989) and the Configuration Hypothesis (Cacciari and Tabossi, 1988). I summarize these explanatory frameworks with some evaluation below. For this thesis, decomposability is crucial, as it will be the primary principle we evaluate. In other words, I ask whether knowing how to decompose idioms allows speakers to use and understand them properly. I also ask what needs to be stored idiosyncratically in order to understand an idiom, and the Configuration Hypothesis focuses upon this latter question.

1.4.1 Decomposability

As we observed by comparing the flexible idioms in (8) and the apparently inflexible idioms in (9), there exist idioms with no apparent syntactic difference that nevertheless exhibit different syntactic behavior. Moreover, even if we wanted to specify flexibility for each idiom in the lexicon, it is difficult to imagine how speakers would learn each idiom’s value for a feature like [\pm Passive] (Nunberg, 1978). Even semantics do not fully explain the distinction. Nunberg (1978) points out that *kick the bucket* and *give up the ghost* both roughly mean ‘to die,’ but *give up the ghost* accepts passivization (10b), while *kick the bucket* does not (10a):

- (10) (Nunberg, 1978)
- a. #Once **the bucket has been kicked**, there is nothing medical science can do.
 - b. Once **the ghost has been given up**, there is nothing medical science can do.

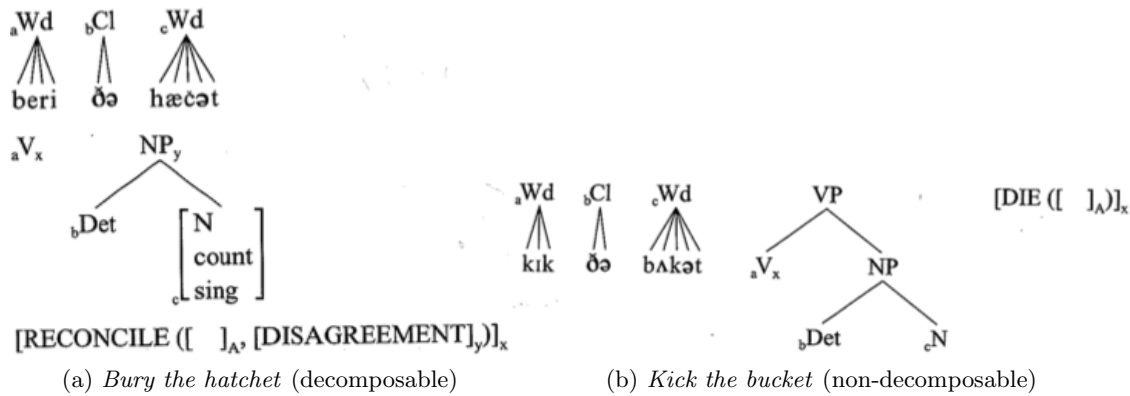


Figure 1.1: Lexical items from Jackendoff (1997)

From this, Nunberg (1978) concludes that it is neither the syntax nor the semantics of an idiom that determines its ability to passivize. Rather, it is the *mapping* between syntax and semantics that produces this effect. Nunberg initially proposes the notion of *decomposability*, which we can summarize as “parts of the idiom’s meaning correspond exactly to parts of the verb.” *Bury the hatchet* is decomposable because *bury* refers to “reconcile,” and *the hatchet* refers to “a disagreement.” Since *kick the bucket* does not have this property, it fails to passivize. Gibbs and Nayak (1989) performed some offline experiments to confirm that decomposability correlates with ability to syntactically modify an idiom. They operationalize the concept of decomposability with their *Idiom Decomposition Hypothesis*: “the syntactic behavior of idioms is determined, to a large extent, [by] speakers’ assumptions about the way in which parts of idioms contribute to their figurative interpretations as a whole” (p. 100).

In contrast with the Configuration Hypothesis, the Idiom Decomposability Hypothesis has received attention in the formal linguistic literature. Most notably, Jackendoff (1995) proposes a phrasal lexicon that includes not only idioms but also other fixed expressions (in the “*Wheel of Fortune* corpus”). In his view, a lexical item is comprised of phonological information, morphological information, syntactic information, and semantic information. Based on this framework, Jackendoff (1997) proposes that decomposable and non-decomposable idioms contain different sorts of lexical entries. For decomposable idioms, the syntactic sub-components and semantic sub-components are all co-indexed, represented in Figure 1.1’s representation of *bury the hatchet* by the subscripts x and y , corresponding to *bury* (‘reconcile’) and *hatchet* (‘disagreement’), respectively. In the case of VP idioms, the entire structure of the VP is not specified; instead, the V and NP are separate components of the representation. For a non-decomposable idiom, on the other hand, the NP in the syntactic representation is not coindexed with any part of the semantic representation, and the structure of the entire VP is pre-specified and therefore frozen.

This approach initially appears like a clean way to reconcile our understanding of the lexicon with our understanding of idiom flexibility. However, it proves problematic when we look beyond idealized examples. To start with, it potentially overpredicts the flexibility of decomposable idioms.

For example, there is no reason that the representation of *bury the hatchet* in Figure 1.1a would preclude a cleft construction. However, cleft constructions sound terrible (11a), even when the equivalent sentence with the non-idiomatic *reconcile the disagreement* sounds perfectly acceptable (11b). There do exist decomposable idioms for which clefting is acceptable, such as *pull strings* (8b) and *draw the line* (12), but at this point, there is not enough documentation in the literature to determine whether they are the exception or the rule.

- (11) a. ?# **The hatchet that they buried** had been tearing their families apart.
 b. **The disagreement that they reconciled** had been tearing their families apart.
- (12) **The line** that Leah’s parents **drew** felt far too restrictive to her.

We might try to save our notion of decomposability by dividing it into further subcategories. Indeed, Nunberg’s (1978) original proposal contained two categories of decomposable idioms: normally decomposable and abnormally decomposable. He distinguishes abnormally decomposable idioms as idioms where “we can identify the referents of their constituent terms, but it is only in virtue of our knowledge of conventional metaphors that we know what that relation is invoked to identify” (p. 228). Under this categorization, it seems possible that *bury the hatchet* constitutes a metaphor whose parts we can distinguish, but which requires invocation of the entire image in the process of decomposability.

Gibbs and Nayak (1989) did find some weak corroboration for this distinction between normal and abnormal decomposability, but the distinction seems to have been abandoned in the subsequent literature. For example, Nunberg himself does not refer to abnormally decomposable idioms at all in Nunberg et al. (1994), choosing instead to introduce the new terminology of “idiomatically combining expressions” and “idiomatic phrases,” which correspond to decomposable and nondecomposable idioms, respectively. Jackendoff (1997) also makes no mention of the distinction and even uses *bury the hatchet* as an example of a decomposable idiom, suggesting that he does not consider the distinction to be important.²

Regardless, the idea of abnormal decomposability was a useful first attempt at capturing an important intuition: decomposability is not actually a matter of binary categories; it is probably best treated as a spectrum. Gibbs and Nayak (1989) actually recognized this, noting that “idiom phrases exist on a continuum of analyzability ranging from those idioms that appear to be highly decomposable (e.g., *pop the question*) to those that can be viewed as semantically nondecomposable

²While I am not aware of any direct explanation for the rejection of “abnormally decomposable” idioms as a category, I can think of several reasons to do so. First, in the case of Jackendoff’s lexical model, it seems unclear how to formally represent the difference between normal and abnormal decomposability; either the syntax of the idiom is coindexed with the parts of its semantic representation or it is not. Second, “normality” of decomposability bears very little explanatory power when we look at more examples of VP idiom flexibility. For example, *bury the hatchet* cannot be used in cleft construction but can passivize, so it is less flexible than a normally decomposable idiom should be but more flexible than an abnormally decomposable idiom should be. Third, the label of “normality” is decidedly *post hoc*, if not meaningless.

(e.g., *kick the bucket*).” However, Jackendoff’s account as stated clearly contradicts this provision.³ Since Gibbs’s work is more psychological and Jackendoff’s is more formal, these contradictions have not been fully fleshed out. However, Espinal and Mateu (2010) provides an interesting formal account showing that the distinction between decomposable idioms (ICEs) and nondecomposable idioms (IPs) proves quite messy.

Many of these general observations calling decomposability to question have been corroborated experimentally. Tabossi et al. (2008) provide a more systematic account of the failures of the idiom decomposability hypothesis (IDH). In a series of experiments, they confirm that subjects’ intuitions regarding decomposability are consistent only for a handful of idealized cases (Experiment 1). This result resembles that of Titone and Connine (1994), who observed consistent decomposability judgments for a minority of the idioms across their 171-idiom norm list. Tabossi et al. (2008) also fail to find evidence that decomposability affects the syntactic flexibility of idioms when they attempt to replicate Gibbs and Nayak (1989) (Experiment 2). Finally, participants are significantly faster at identifying idiomatic strings as valid Italian phrases than matched control strings, but Tabossi et al. fail to find any effect of decomposability or even flexibility. While we need to be careful about drawing strong conclusions from null results, it is clear from our examination of example sentences and from studies such as this that the strongest predictions of IDH are not borne out in the data. Furthermore, we might be disinclined to encode decomposability in the lexicon if it never proves significant in on-line studies.⁴

Overall, IDH provides a governing principle that could explain what speakers need to know in order to correctly determine when an idiom can be flexible. However, an initial closer examination of the data calls to question its explanatory power. I will use IDH as a starting point for understanding the flexibility of idioms throughout this thesis, and both of my experiments are designed to test its explanatory power and determine how to expand its explanatory power.

1.4.2 Configuration Hypothesis

Unlike IDH, the Configuration Hypothesis says very little about the source of idiom flexibility; rather, it provides an account of the other important aspect of speakers’ knowledge of idioms: idioms’ representation in the mind. The most central tenet of the Configuration Hypothesis is that idioms are not, in fact, lexical items. Earlier lexical models, such as the Direct Access Hypothesis (Gibbs, 1980) and the Lexical Representation Hypothesis (Swinney, 1979), would respectively

³It might be saved by coindexing the semantics and syntax in subtly different ways, but at this point we are examining the issues as they stand, not their resolutions.

⁴To my knowledge, only two on-line psychological experiments have ever found an effect of decomposability. Titone and Connine (1999) found a weak effect of contextual reading speed for non-decomposable idioms, while Caillies and Butcher (2007) found the decomposable idioms in French are processed sooner than non-decomposable idioms, based on a Cross-Modal Lexical Priming design. These results, faced with the general lack of evidence, might be better understood in terms of idiom transparency (the clarity of the underlying metaphor), according to Tabossi et al. (2008).

require the idiomatic interpretation to precede or run in parallel to the literal interpretation. However, Cacciari and Tabossi (1988) found that the idiomatic meaning for $[_{VP} V (NP) (PP) (PP)]$ idioms creates no facilitation effect until the final word of the idiom, and this effect is only observed for predictable idioms. In the case of non-predictable idioms, only the last word's literal meaning produced a facilitation effect. Idiomatic activation occurred at a 300ms delay, but the literal meaning of the final word remained activated.

Because of this, Cacciari and Tabossi propose that idioms are “configurations of words” whose meaning is accessed “whenever sufficient input has rendered the configuration recognizable.” They consider this to be analogous to Marslen-Wilson’s (1987) cohort model for word recognition, and they propose that idioms contain *keys* whose detection enables idiom recognition (analogous to Marslen-Wilson’s “uniqueness point”). Over the years, this model has been developed more schematically, incorporating results from production models such as the “Superlemma hypothesis” of Cutting and Bock (1997) and Sprenger et al. (2006). In this model, the lexicon is comprised of a conceptual level, where meaning is represented; a lexical-conceptual level, where the nodes connecting form and meaning are found; and a lexical-syntactic level, where each word’s syntactic category is stored. A “configuration” is not a node, but “a co-occurrence of a series of nodes corresponding to the multi-word expression we have learned” (Tabossi et al., 2009b). Tabossi et al. point out that this model easily accommodates other formulaic expressions; idioms differ only in the presence of a unique conceptual level. Tabossi et al.’s (2009b) diagram of the lexicon under this model has been replicated as Figure 1.2.

1.4.3 Comparing the Models

Comparison of both of these models allows us to see what is actually agreed upon regarding what speakers need to know about idioms in order to determine their flexibility. While Tabossi and colleagues are highly critical of the Idiom Decomposition Hypothesis (Tabossi et al., 2008, 2009a), actual comparison of the Configuration Hypothesis to IDH proves challenging. The CH is a model of psychological representation of multi-word expressions, while IDH, as originally stated by Gibbs and Nayak (1989), only comments on the relationship between semantic decomposability and syntactic flexibility. By itself, IDH is fairly agnostic regarding the actual representation of idioms, and CH is fairly agnostic regarding flexibility, with the exception of asserting that it is not lexically specified (Tabossi et al., 2009b).

The more worthwhile comparison, then, is between CH and Jackendoff’s representation of the lexicon based on decomposability. In other words, how do Figures 1.1a and 1.2 differ? Most saliently, Jackendoff writes from more of a formal perspective, while CH is a psychological model, leading to superficially different diagrams. Moving past these superficial differences, however, we can observe several important theoretical distinctions.

First, Jackendoff believes that the representation of each idiom contains a representation of the

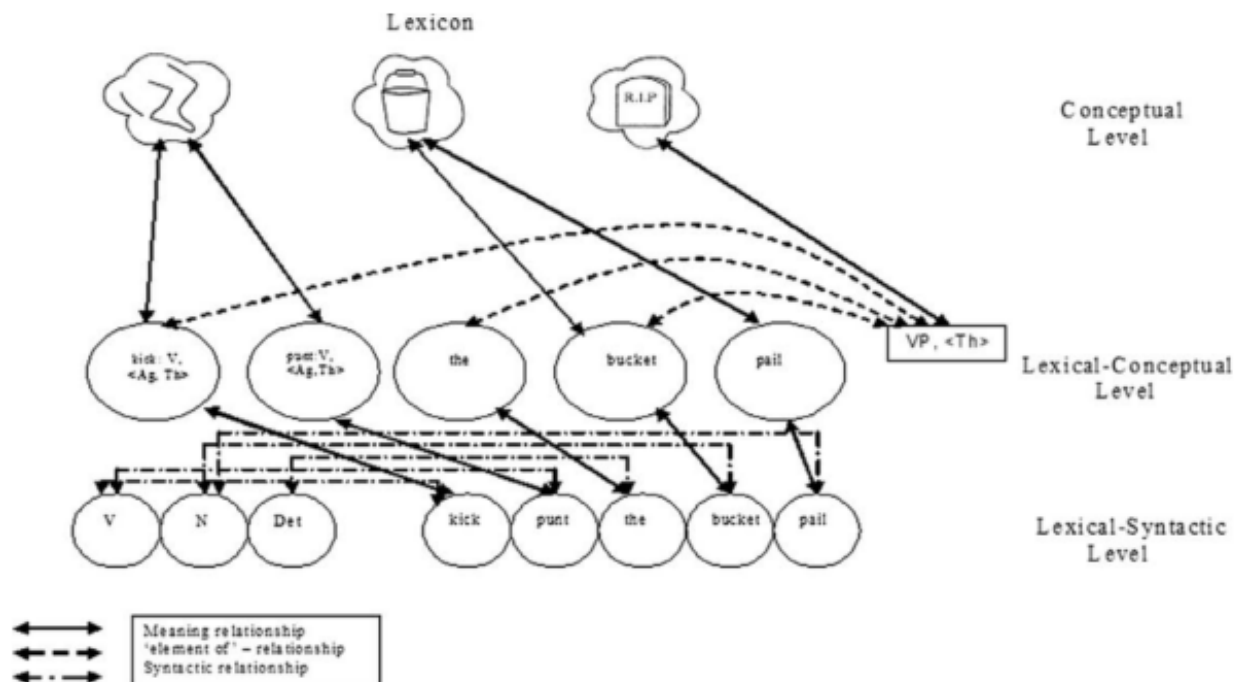


Figure 1.2: Representation of *kick the bucket* according to the Configuration Hypothesis, from Tabossi et al. (2009b).

phrase's syntax, while Tabossi et al. (2009b) explicitly reject this. Tabossi and colleagues reason that flexibility varies considerably based on context and speaker. For example, *speak one's mind* rarely accepts adjective insertion, but adjective insertion can be acceptable in sentences like (13). From this, they conclude that lexicalized syntax cannot explain the variety observed in the data.

- (13) When drugs are involved, it's time to **speak your parental mind**. (Glucksberg, 2001)

However, Tabossi et al. miss an important distinction between lexicalized *syntax* and lexicalized syntactic *flexibility*. The data quite clearly demonstrate that the inclusion of a [\pm adjective-insertion] feature would not be useful (indeed, this observation dates back to Nunberg (1978)), but Jackendoff would likely argue that the syntax comes "for free" by virtue of representing both the form and meaning of the idiom with its constituents. In other words, if we were to collapse Figures 1.1a and 1.1b such that decomposability were not encoded, Jackendoff would not truly contradict the claims of the CH model; rather, the contradiction lies in the understanding of what lexical syntax is about.

Second, Jackendoff believes that stored form-meaning combinations are by definition lexical, while a central tenet of CH is that idioms are "configurations" that do not constitute lexical items, i.e. they do not have lexical-conceptual nodes. Jackendoff (1995) summarizes: "From the point of view of modularity, what counts as part of language is anything stored in memory in phonological and/or syntactic format." We definitely do not want to exclude formulaic language from our understanding

of the linguistic system, but this still does not answer the question of whether idioms need to be in the lexicon to be part of the linguistic system.

I do not think that proponents of CH are suggesting that idiomaticity lies outside of the linguistic system. Rather, they question what it means to be a lexical item. Their argument seems to reduce to the fact that idioms do not function exactly like words, and since words must be lexical items, idioms must not be. Differently put, idioms contain lexical items, so they cannot be lexical items themselves. This conclusion does not seem necessary, especially if we allow that non-word morphemes have lexical entries, since these require some redundancy in the lexicon, with some particularities specially encoded (Jackendoff, 1995). Regardless, the primary difference here seems to be one of *definition* of a lexical item, not a major representational question.

Third, the sense of connection between the words in a given idiom and those same words alone differs, and this is likely the source of the second distinction. In CH, the connection between words the words in the idiom's representation and the words as standalone lexical items is clear; the idiom is simply a configuration of those standalone lexical items. This not only counts the bounty of psychological data showing priming for literal meanings of words; it captures the important observation that the words in idioms share irregular morphology with their standalone counterpart (14):

- (14) a. We **shot** the breeze for a while yesterday.
 b. *We **shoote**d the breeze for a while yesterday.
 c. *We [**shoot the breeze**]d for a while yesterday.

Jackendoff (1995), on the other hand, expresses a willingness for the existence of complete lexical redundancy between the internal words in idioms and those words as standalone lexical item, citing the accepted and existing need for such redundancy in morphological paradigms. He later explains this in terms of semantic inheritance networks (Jackendoff, 2002). This question of the relationship between standalone words as lexical items and those words stored in the context of an idiom remains open, but it is not the focus of the present study.

To summarize the popular models, idiom decomposability offers a partial explanation of idiom flexibility by appealing to the mapping between syntactic and semantic structure within an idiom, and the Configuration Hypothesis offers a psychological model of how idioms are stored. Despite receiving support from opposing camps in the research literature on idioms, it seems that the two models differ more in what they seek to explain than they actually contradict each other. Therefore, we can seek to understand their findings in one coherent model rather than pitting them against each other for sociological reasons. Important to the questions at hand, both models are consistent with each idiom being stored in some mental unit (regardless of its label); they just differ on whether syntactic flexibility is a property inherent to this mental unit.

1.5 Open Issues

By understanding how idioms are represented and what we need to know about them to treat them with appropriate flexibility, we will begin to answer the question of what speakers need to know in order to properly use an idiom. We now have an overview of the phenomena of idiomaticity and idiom flexibility, with a sense of some of the representational models that have been adopted. Despite all of the work that has been done since Nunberg (1978) first proposed the concept of decomposability, very little consensus exists regarding any explanation for the details of idioms' representation in the mind and any other principles to explain. We only know that decomposability at best explains a fragment of the data. What are the rules, how many are needed, and is anything idiom-specific? Decomposability and the Configuration Hypothesis have started to ask these questions, but no clean answers exist.⁵

Beyond this, the existing data are quite limited. All of the existing discussion on idiom flexibility looks only at VP idioms, often restricting the discussion to [V NP] idioms. However, there exist plenty of NP (15), PP (16), and AP (17) idioms. Therefore, we risk falling to the same trap as Chomsky (1981), looking at a nonrepresentative set of examples at the expense of marginalizing others and missing important generalizations.

(15) can of worms, piece of cake, fish out of water, two peas in a pod, par for the course

(16) out of the blue, in hot water, with flying colors, under the weather

(17) dressed to kill, fit as a fiddle

Taking this into account, we will proceed to account for speakers' ability to accept certain types of flexibility for certain idioms by understanding what needs to be placed in the lexicon and where to look for overarching principles. Chapter 2 presents the first original experiment, which asks whether flexibility needs to be stored idiosyncratically in the lexicon, as well as determining the explanatory power of several factors such as decomposability when it comes to idiom flexibility. This study also provides an expansive data set for postulating further principles and extensions of idiom flexibility. Chapter 3 further tests the boundaries of decomposability by presenting an original experiment that addresses a confound in Gibbs and Nayak's (1989) original study. This study will prove germane to the general question by qualifying IDH and by determining the reliability of speakers' conceptualization of idioms. Finally, in Chapter 4 we will examine the data from both experiments, determining both the principles explaining patterns of idiom flexibility and how these relate to the lexicon.

⁵We also do not have a definite picture of the relationship in the lexicon between words in an idiom and their standalone counterparts. This question is important and even relevant to our discussion of flexibility, but for the sake of space, it will not be our focus in the present thesis.

Chapter 2

Experiment 1: Opening the Can of Worms

2.1 Background and Purpose

We are trying to uncover the source of speakers' idiom flexibility judgments. One potential explanation for idiom flexibility is that speakers store idiosyncratic information about each idiom's syntax in the mind, perhaps in the mental lexicon. Sprenger et al. (2006), for example, propose a "superlemma hypothesis," where each idiom's syntax is part of its mental representation.¹ Under this interpretation of the superlemma model, to know how to properly use an idiom means to have the correct mental representation of the idiom in the form of a superlemma. Tabossi et al. (2009b) refute this by showing that speakers have equally reliable intuitions regarding the flexibility of real idioms as they do for idioms that are either made up or unfamiliar. If competent judgments for idiom flexibility required access to prespecified lexical syntactic flexibility, we would have expected greater agreement among speakers regarding real idioms than regarding novel idioms.

From this, Tabossi et al. (2009b) conclude that idiom syntax is principled, not idiosyncratic; idiom syntactic flexibility does not require idiosyncratic knowledge of the idiom. The present study expands on their study by addressing two concerns. First, was Tabossi et al.'s (2009b) result simply a rest of an idealized set of data and an overly coarse-grained measurement (participants gave binary acceptability judgments rather than using a larger scale)? Second, Tabossi et al. suggest that idiom syntax must be completely based on "principles"; in order to use an idiom properly, a speaker must have internalized a series of principles that can apply to idiom usage. But what are these? This study is designed to answer both of those questions. I address issues of representativeness by selecting a group of frequently-attested idioms of diverse syntactic categories

¹This does not necessarily mean that all syntactic flexibility is pre-stored in their model, but this is one interpretation of their model, which Tabossi et al. (2009b) take to be the case.

(rather than only VP idioms), and I address the issue of granularity by replacing the binary scale with a 1-7 (Likert-type) scale. On this level, the present study allows us to confirm that Tabossi et al.'s (2009b) results truly represent idioms and that their effect was not washed out by the resolution of their data. This endeavor lets us finally ask the question of whether speakers need to know specific flexibility information about idioms in order to properly use them.

If flexibility is not idiosyncratically stored as part of an idiom's lexical representation,² then speakers must know some principles in order to determine how to acceptably use idioms. This study is designed to point toward potential principles underlying idiom flexibility. It led to the collection of a large paradigm of sentence judgment data to serve as the basis of theorizing, allowing for item-based analyses (see Chapter 4), but it also allowed for linear modeling to determine the degree to which factors such as decomposability are predictive of idiom flexibility. Beyond these broad-strokes analyses to lead to some general understanding of principles explaining idiom flexibility, this study examined the possibility that the principle of scalar substitution could predict some idiom flexibility. Before turning to the specific design of the study, I will explain what I mean by scalar substitution.

2.2 Scalar Substitution

I wish to systematically validate the existence of a particular type of word substitution for idioms, which I call *scalar substitution*. I already alluded to this in the introduction when presenting some naturally occurring examples, replicated in (18). These examples involve the production of a novel meaning by virtue of comparison of the original word *top* to the substituted word *middle*, and evaluating their difference on a scale. McGlone et al. (1994) suggested this with a broader appeal to pragmatics with the example replicated in (19a). I suggest the importance of a clear scale to rule out substitution of words with other relationships to the original, as illustrated by the unavailability of (19b). Unlike *nudge*, which we would characterize as a type of caused movement that is less than kicking, *punting* is a type of *kicking*. Thus, (19b) is unavailable because the substituted word has a proper subset relationship to the original word rather than a scalar relationship.

(18) It's not worth fucking your way to the **middle**. (*House of Cards*, E9)

(19) a. Tom: *Did the old man kick the bucket last night?*
 Joe: *Nah, he barely **nudged** it* (McGlone et al., 1994)

b. ... Joe: #Nah, he **punted** it.

Importantly, scalar substitution is a productive type of idiom modification that does not depend on decomposability; it is possible for *kick the bucket*, the most prototypical of nondecomposable/inflexible idioms. Additionally, it is particularly useful in an experimental setting involving

²Or "configuration"

non-VP idioms, since, unlike passivization, clefting, topicalization, etc., it does not require that an idiom be a VP in order to be logically possible. Before I commence with testing my intuitions empirically, I should address an immediate objection to scalar substitution: it feels like metalinguistic wordplay, rather than part of the linguistic system. This is not necessarily true, since McGlone et al. (1994) found that subjects process variant idioms at the same rate as literal controls, which are both slower than idioms; we would expect a metalinguistic process to produce a processing delay. However, even allowing for the possibility that this phenomenon is metalinguistic, it still provides useful clues about the contribution of parts of idiomatic expressions to the overall meaning, as the pattern in (19) shows us.

2.3 Design and Methodology

2.3.1 General Design

Tabossi et al.'s (2009b) Experiment 1 has been modified in several crucial ways. First, the materials have been expanded to not only include VP idioms, but also NP idioms and PP idioms. Second, the measurement of acceptability has been changed from a binary judgment to a 1-7 (Likert-type) scale. Third, the analysis of subject agreement is based on standard deviation values rather than percentage of subjects who agree with the majority, in order to allow any smaller distinctions to become manifest.

2.3.2 Amazon Mechanical Turk

The survey was distributed using the crowdsourcing mechanism Amazon Mechanical Turk (henceforth, “MTurk”). This platform allows for the recruitment of respondents to complete “Human Intelligence Tasks,” generally tasks that are simple for a human but difficult to accomplish with a computer program. This platform has become increasingly popular for survey-based work in recent years due to its low cost (Paolacci et al., 2010), fast completion time, and diverse participant pool (Behrend et al., 2011). The validity of its data has been confirmed by several psychological studies (Mason and Suri, 2012; Behrend et al., 2011), and to a lesser extent, for linguistic work (Munro et al., 2010).

I determined that MTurk was the best choice for this project because of the large amount of data I was able to collect in a short time, permitting follow-up, but more importantly, the diverse participant pool is very informative here. It is quite possible, if not likely, that different idioms show different familiarity across age groups and regional groups. Since I want to understand idiomaticity in American English as a whole, the use of MTurk provides a more representative sample than would a group of Yale undergraduates.

2.3.3 Participants

A group of 200 participants were recruited through Amazon Mechanical Turk. They were each paid \$1.00 for an average of 25 minutes spent completing the survey. Regardless of their performance, their HIT was approved; even participants whose data failed to meet my quality standards were paid. Only MTurk workers who had completed a minimum of 1000 hits with a 95% acceptance rate or above were granted access to the task, which was also restricted to IP addresses in the United States.

By participating in the survey, participants claimed to be native speakers of American English. No further biographical data was gathered, although other studies have shown the demographics of MTurk to be diverse in age and skewed toward females (Behrend et al., 2011).

2.3.4 Materials

Idioms

For the survey, I sought a set of idioms to be maximally representative of the diversity of idiomatic expressions, in addition to comparing both real and novel idioms. (As a reminder, real and novel idioms were used to determine whether foreknowledge of an idiom leads to more reliable flexibility judgments, a proxy measure for lexical encoding.) To this end, three different syntactic categories were selected, leading to the inclusion of 15 idioms that constitute a verb phrase (VP idioms), 10 idioms that constitute a noun phrase (NP idioms), and 10 idioms that constitute a prepositional phrase (PP idioms). More VP idioms were included than NP or PP idioms to allow for more tokens of the passive manipulation, which is not possible for NP and PP idioms. The real idioms were taken from the Titone and Connine (1994) norms. In order to be maximally representative of actual idiom usage, the most frequent idioms of a given syntactic category were selected. A complete list of the idioms used is presented in (20), and their meanings can be found in Appendix A.

(20) Real Idioms

- a. **VP:** *to give someone the cold shoulder, to slip someone's mind, to get the picture, to call the shots, to face the music, to hit the sack, to swallow one's pride, to jump the gun, to learn by heart, to twist someone's arm*
- b. **NP:** *food for thought, rule of thumb, tip of the iceberg, a frog in one's throat, fish out of water, two peas in a pod, par for the course, letter of the law*
- c. **PP:** *out of the blue, out of thin air, over the hill, in hot water, in the nick of time, with flying colors, under the weather, in seventh heaven, at the back of one's mind, on thin ice*

An equally distributed set of novel idioms was also selected: 15 VP, 10 NP, and 10 PP. These “novel” idioms were adapted from calques of Finnish and Swedish idioms. These particular languages were selected because they are highly unlikely to be spoken by native speakers of American English, but as European languages, their cultural context should be compatible with that of American English speakers. Moreover, I chose to use true idioms of other languages rather than generating fake idioms (as done by Tabossi et al. (2009b)) for two reasons. First, these idioms are less likely to contain a metaphorical bias; that is, idioms actively generated by a person will be more likely to transparently have a basis in a metaphor. Second, the goal of the study is to understand *novel* idiom behavior – how participants judge idioms they have never heard before. While the ideal would be idioms of other dialects of English (cf. Tabossi et al. (2009b), Experiment 2)), such a long and thorough list is impossible to find. Thus, idioms from other languages are the closest possible way to approximate how subjects would understand actual idioms that they simply had never heard before.

In order to reach 10 PP idioms, a few alterations were necessary. *Dropped behind the wagon* (‘stupid’) was changed to *on the wagon* (‘intelligent, put together’); *hot on the porridge* (‘very eager’) was changed to *on the hot porridge* (meaning preserved); and *curious in a cone* (‘very nosy’) was changed to *in a curious cone* (meaning preserved). The complete list of novel idioms used is found below in (21); their meanings can be found in Appendix A.

(21) Novel Idioms

- a. **VP:** *to roll hat, to walk on the pump, to throw the spoon, to drop the suck, to hit a drill, to walk in the roof, to grasp the gallop, to give someone the rake, to twist from wire, to put on someone’s fallen cloak, to begin on a new marble, to play one’s mouth, to throw a goat’s eye, to speak out of the beard, to put onion on the salmon*
- b. **NP:** *ice in the stomach, fly leg in the universe, someone’s heavy coin, cake on top of cake, a bridge of a donkey, a chicken to pluck, a dog buried somewhere, snows of past winters, brother-in-law politics, nose-bleaching*
- c. **PP:** *behind the floating, on the wallpaper, in the count’s time, in dark clouds, on the hot porridge, on the wagon, against the walls, with helmet askew, with one’s beard in the mailbox, in a curious cone*

A summary of all of the materials can be found in Table 2.1.

Table 2.1: Summary of Materials

	VP	NP	PP
Real	15	10	10
Novel	15	10	10

Modifications

Each idiom was presented in one of five possible forms (“modifications”): control, adjective-inserted, scalar substituted, and non-scalar substituted. The *control* form was simply the canonical/citation form of the idiom (22a). For the *adjective inserted* form, a contextually appropriate adjective was inserted before a noun within the idiom (22b). In the case of *scalar substitution* (22c), one word in the idiom was substituted with another word that can be placed on some scale relative to the original word. By virtue of the comparison between the two words, a new meaning is produced (see previous discussion on *nudge the bucket*). As a counterpart to scalar substitution, *non-scalar substitution* (22d) involved word substitution of a similar word, roughly a synonym, such that the comparison between the new word and the original does not produce any interesting change in meaning.

These four modification conditions were selected for two reasons. First, unlike well-documented phenomena like passivization and topicalization, they are not restricted by the idiom’s syntactic type. Second, very limited experimental work has examined word substitutions, so the inclusion of these conditions contributes new knowledge of what modifications are even possible. In the case of VP idioms, a fifth condition was included: *passivization*. Each idiom was placed in passive form, with the subject placed in a *by*-phrase. In the case of VP idioms that take the form [_{VP} VP PP], the passivization involved movement of the NP in the PP to subject position, leaving behind the preposition (23). All conditions were created by starting with a control sentence and modifying it minimally to produce the other condition, as illustrated by the paradigm in (22). While this definitively eliminated the option of a within-subjects design, it has the benefit of counterbalancing across conditions.

(22) *to break the ice* (‘to start a conversation’)

- a. **Control:** At the awkward cocktail party, Harry told a great joke in order to **break the ice**.³
- b. **Adjective Insertion:** At the awkward cocktail party, Harry told a great joke in order to **break the stuffy ice**.
- c. **Scalar Substitution:** At the awkward cocktail party, Harry told a great joke, totally changing the atmosphere and **shattering the ice**.
- d. **Non-scalar Substitution:** At the awkward cocktail party, Harry told a great joke in order to **split the ice**.
- e. **Passive:** At the awkward cocktail party, Harry told a great joke, and **the ice was broken by him**.

(23) When Joe joined a gang in order to feel protected, **fire was being played with** by him.

³Boldface is for illustration purposes only; participants saw sentences without any boldface.

Experimental Lists

In order to ensure that each participant saw a given idiom only once, it was necessary to create five experimental lists (each idiom-containing sentence appeared in a maximum of five versions experiment-wide). Five lists were created where each idiom was represented at most once, and all conditions were evenly represented, as summarized in Table 2.2. Since NP idioms and PP idioms were presented in only four manipulations, each NP and PP idiom was missing from exactly one of the five experimental lists. This means that a given list contained 16 of the 20 NP idioms and 16 of the 20 PP idioms, omitting 2 real and 2 novel idioms in both cases. Therefore, each list contained 62 sentences. Each subject was presented with one of these five lists.

Table 2.2: Item representation in each list – the number of items of each type seen by a given participant

	Control	Adj Insertion	Scalar Sub.	Non-Scalar Sub.	Passive
Real	3 VP, 2 NP, 2 PP	3 VP, 2 NP, 2 PP	3 VP, 2 NP, 2 PP	3 VP, 2 NP, 2 PP	3 VP
Novel	3 VP, 2 NP, 2 PP	3 VP, 2 NP, 2 PP	3 VP, 2 NP, 2 PP	3 VP, 2 NP, 2 PP	3 VP

2.3.5 Procedure

After selecting the HIT on Amazon Mechanical Turk, participants were directed to a link to a survey made with and hosted by the Qualtrics survey software. They were presented with a welcome screen explaining the length and demand of the task, followed by a second screen that more thoroughly explained their task. First, they were instructed that they would see an idiom/meaning. Then, they were told to rate their familiarity with the idiom, as replicated below:

In the following pages you will be presented with an idiom (e.g. *kick the bucket*) and its meaning (e.g. ‘to die’). You will be asked to rate your familiarity for each idiom on a scale from 1-7.

Select “1” if you have never heard the idiom before. Select “7” if you hear and/or use it all the time, and select an appropriate number in between 1 and 7 to indicate different degrees of familiarity (nearer to 1 = less familiar, nearer to 7 = more familiar).

Note that some of the idioms are not real idioms of English, in which case you should select “1.”

After this, participants were given instructions on giving acceptability judgments for the sentential examples of modified idioms. Since the specific wording is particularly important for a task of this nature, the exact text that participants saw is replicated below:

Next, you will see a sentence containing the idiom. Keeping in mind that the idiom might be modified, rate on a scale from 1 to 7 how much sense the usage of the idiom makes.

Select “7” if the sentence is a perfectly acceptable and understandable way to use the idiom. Select “1” if the idiomatic meaning of the phrase would not make sense in the sentence, and select an appropriate number in between 1 and 7 to indicate different degrees of understandability (nearer to 1 = makes less sense, nearer to 7 = makes more sense).

Participants were reminded in the directions to provide a “gut response” rather than searching on the Internet for a “correct” answer. The instructions were followed by a single example using *kick the bucket*, a popular example idiom but one which did not appear in the experimental materials.

After the instructions, participants were presented with one of the five 62-idiom experimental lists, with the idioms presented in random order for each participant. List assignment was random, but weighted so that each list would be represented equally. For each idiom, participants saw the idiom, its meaning, and a sentence containing one of the four modification conditions. They rated their familiarity with each idiom on a 1-7 scale, then the degree to which the modified version of the idiom made sense on a 1-7 scale. A 1-7 scale was chosen in order to allow for participants to show a wide range of judgments. Particularly, we wanted participants to be able to consider a sentence to be more or less acceptable, but not completely so, illustrating the difference between “5” and “6.” In the case of familiarity, the wider scale allowed participants both to categorize the idioms as real or novel and to show their level of familiarity with the real idioms.

At the end of the survey, participants were given a chance to leave comments, and then they were given a confirmation code to copy into the Amazon Mechanical Turk HIT interface.

2.4 Results

In analyzing the large set of data produced by this experiment, I performed several distinct analyses, described below. First, I excluded participants whose data was of insufficient quality (Section 2.4.1). After this, I examined their judgments for each condition, notably finding a general preference for real idioms and a general dispreference for scalar substitution (“shatter the ice,” Section 2.4.2). The primary purpose of this analysis is to see what general principles are at play. Next, in order to determine if any idiom’s syntax must be idiosyncratically stored for proper use, I analyze speakers’ consistency in their judgments, observing a difference between real and novel idioms driven by participants’ general agreement on real controls (Section 2.4.3). Finally, I examine factors about idioms such as decomposability and determine their explanatory value for idiom flexibility (Section 2.4.4), concluding with a discussion of some interesting items (Section 2.4.5). All statistical analysis was performed using the R statistical programming language (R Development Core Team, 2011).

2.4.1 Quality Control

Due to the nature of crowdsourced research, it is particularly important to eliminate the data of participants who simply clicked quickly in order to earn their money. Participants were deemed unreliable based on two criteria: false familiarity and atypical completion time. By “false familiarity,” I mean that participants gave a familiarity rating of 3 or higher to novel idioms, thus claiming to recognize items that they could not truly have recognized.⁴ A participant’s data were eliminated if s/he had more than two false positives out of 31 novel idioms, requiring a 93.5% accuracy rating or above. This eliminated 62 participants.

A participant’s data were also eliminated if survey completion time was more than 1.5 standard deviations from the mean (24.4 minutes, 1 *S.D.* = 11.3 minutes). This eliminated 21 participants, 10 of whom also showed too many false positives. Note that this step eliminated more responses from participants who took too long to complete the survey than from those who completed the survey too quickly. However, all four of the participants who completed the survey in under 10 minutes were already eliminated due to their high number of false positives. No correlation was observed between number of false positives and completion time ($F(1, 198) = 0.02489$, $p = 0.8748$).

In total, 73 participants were eliminated from further analysis, leaving 127 whose data were analyzed. Their distribution across the five lists is provided in Table 2.3. While list representation is not equal, there was no significant correlation between list and whether the participant’s data was deemed acceptable ($F(1, 198) = 1.079$, $p = 0.3$).

Table 2.3: Analyzed Participants by List

List	1	2	3	4	5
# of Participants	28	24	29	22	24

All HITs were approved regardless of whether participant data was deemed useable. Additionally, the items *lose one’s grip*, *letter of the law*, and *to give someone the rake* were eliminated due to presentation errors in the experiment.

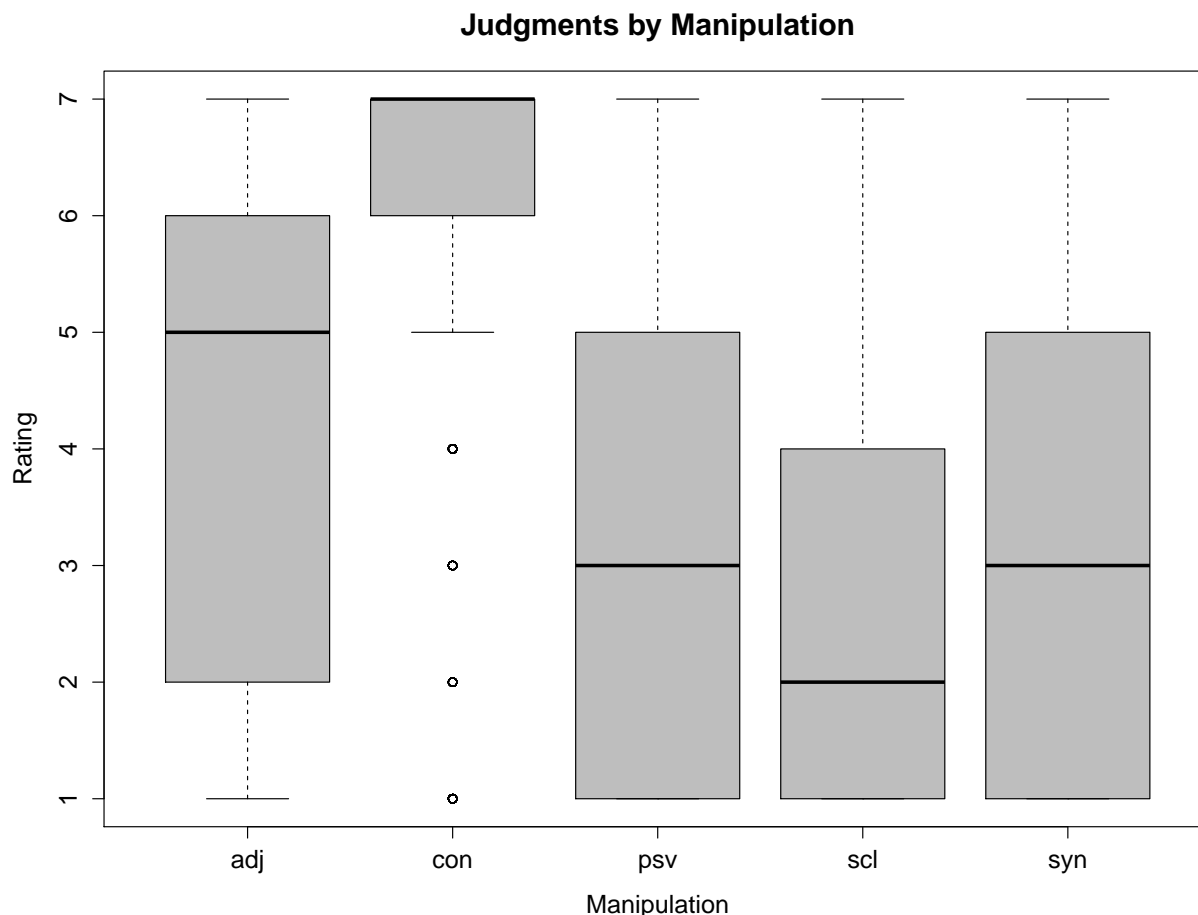
2.4.2 Judgments

With unreliable data excluded, we can now turn to participants’ judgments regarding idioms in different manipulations. Most importantly, we want to see whether scalar substitution (“shatter the ice”) is generally acceptable and therefore an example of a principle speakers use in determining acceptable flexible uses of idioms. Additionally, we want to confirm that the control sentences were generally acceptable.

There was a significant main effect of type of idiom modification ($F(4, 7512) = 690.6$, $p \approx 0.000$).

⁴Alternatively, they could be speakers of Finnish or Swedish, but this would still merit exclusion from analysis.

Figure 2.1: Judgments of manipulation meaningfulness, including both real and novel idioms. Key to conditions: *con*=control (“break the ice”); *adj*=adjective insertion (“break the stuffy ice”); *scl* = scalar substitution (“shatter the ice”); *syn* = non-scalar/synonym substitution (“split the ice”); *psv* = passivization (“the ice was broken”).

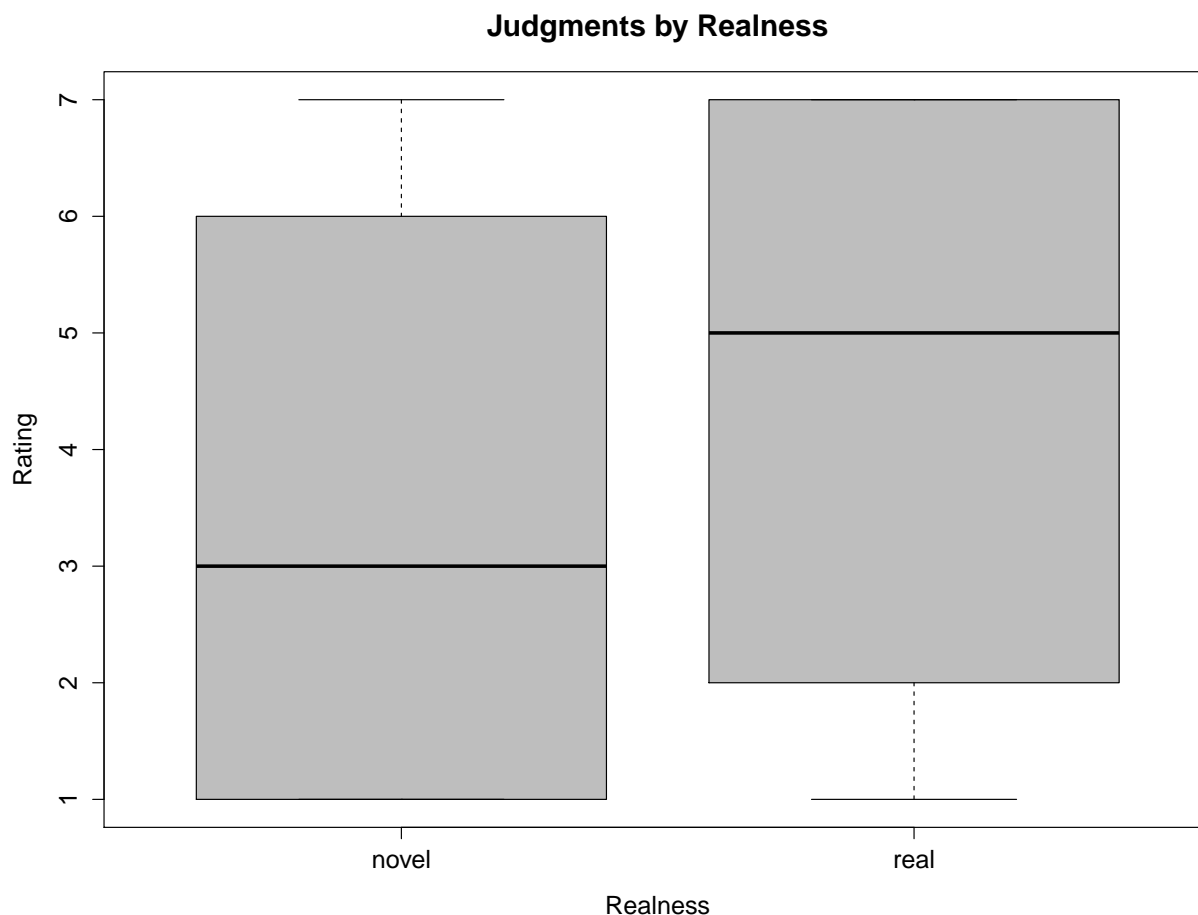


Post-hoc tests reveal that the control condition (“break the ice,”⁵ $M = 6.098$) was rated significantly more acceptable than adjective insertion (“break the stuffy ice,” $M = 4.27$, $p \approx 0.000$); adjective insertion was significantly more acceptable than both passivization (“the ice was broken”, $M = 3.33$, $p \approx 0.000$) and non-scalar substitution (“split the ice”, $M = 3.318$, $p \approx 0.000$); passivization and non-scalar substitution showed no significant difference in acceptability ($p = 1$); and scalar substitution (“shatter the ice”, $M = 2.783$) was rated least acceptable in comparison to all other manipulations (in all cases, $p \approx 0.000$). Despite the significance of these trends, we should note that there was considerable variance within each manipulation, visualized as a boxplot in Figure 2.1.

Beyond this, there was a significant main effect of realness (English idiom vs. novel idiom, $F(1, 7515) =$

⁵I use *break the ice* for the purpose of illustration of conditions; it did not necessarily pattern according to the general trend.

Figure 2.2: Judgments according to realness, including all manipulations.



274.2, $p \approx 0.000$), which can be visualized in Figure 2.2. Participants generally ranked manipulations containing real idioms ($M = 4.489$, $S.D. = 2.289$) as more acceptable than manipulations containing novel idioms ($M = 3.604$, $S.D. = 2.339$). Finally, the two-way interaction between realness and manipulation was significant ($F(9, 7507) = 377.4$, $p \approx 0.000$). *Post hoc* tests reveal that real idioms receive significantly higher acceptability ratings than novel idioms for all manipulations. Moreover, for real idioms, there was a significant hierarchy of acceptability: control was most acceptable, followed by adjective insertion, followed by passivization, followed by non-scalar substitution, followed by scalar substitution. For novel idioms, the hierarchy differed. The control was most acceptable, followed by adjective insertion, followed by non-scalar substitution, with passivization and scalar substitution considered least acceptable, but not significantly different from each other ($p = 1.0$). While these patterns by realness are interesting, they could merely reflect the particular items selected, so they ultimately say very little about the questions at hand.

Figure 2.3: Interactions Plot for Realness and Manipulation. Key to conditions: *con*=control (“break the ice”); *adj*=adjective insertion (“break the stuffy ice”); *scl* = scalar substitution (“shatter the ice”); *syn* = non-scalar/synonym substitution (“split the ice”); *psv* = passivization (“the ice was broken”).

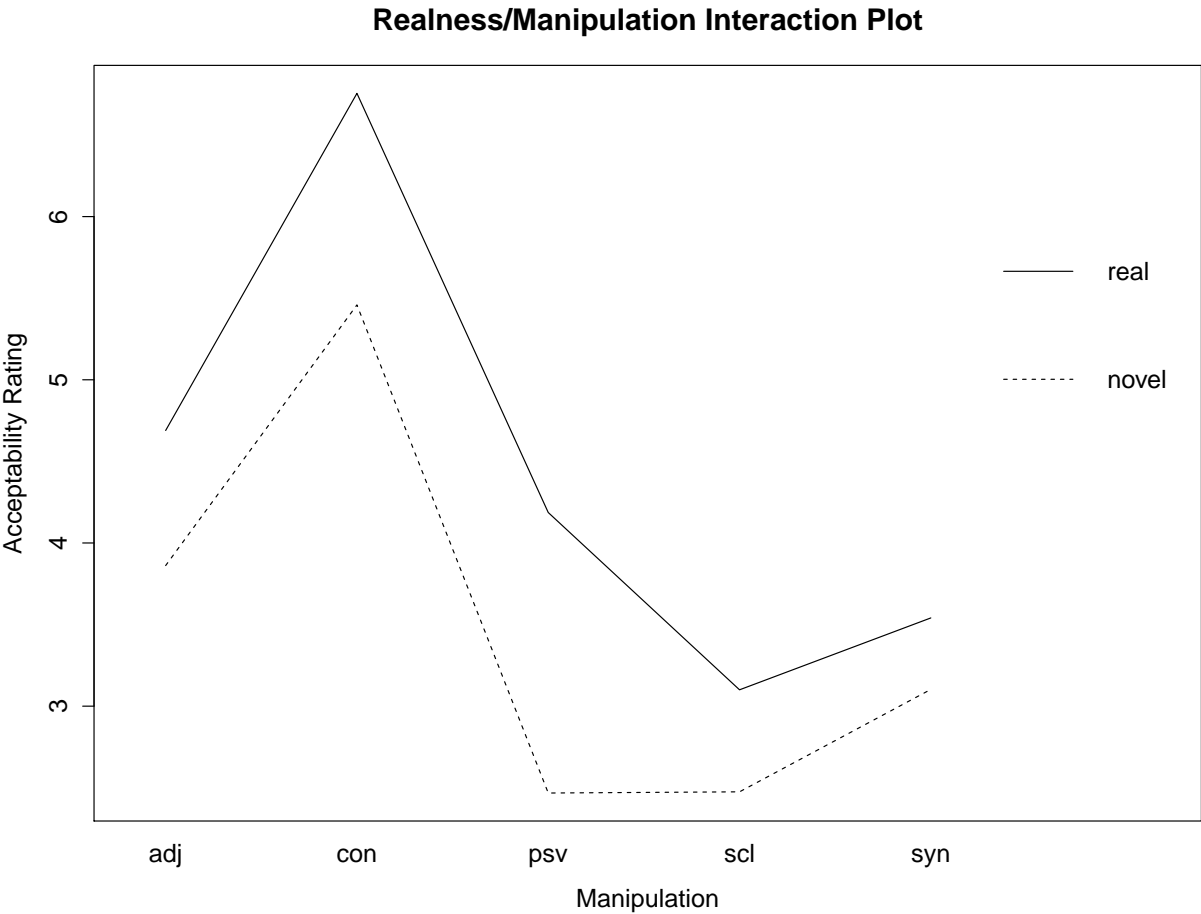
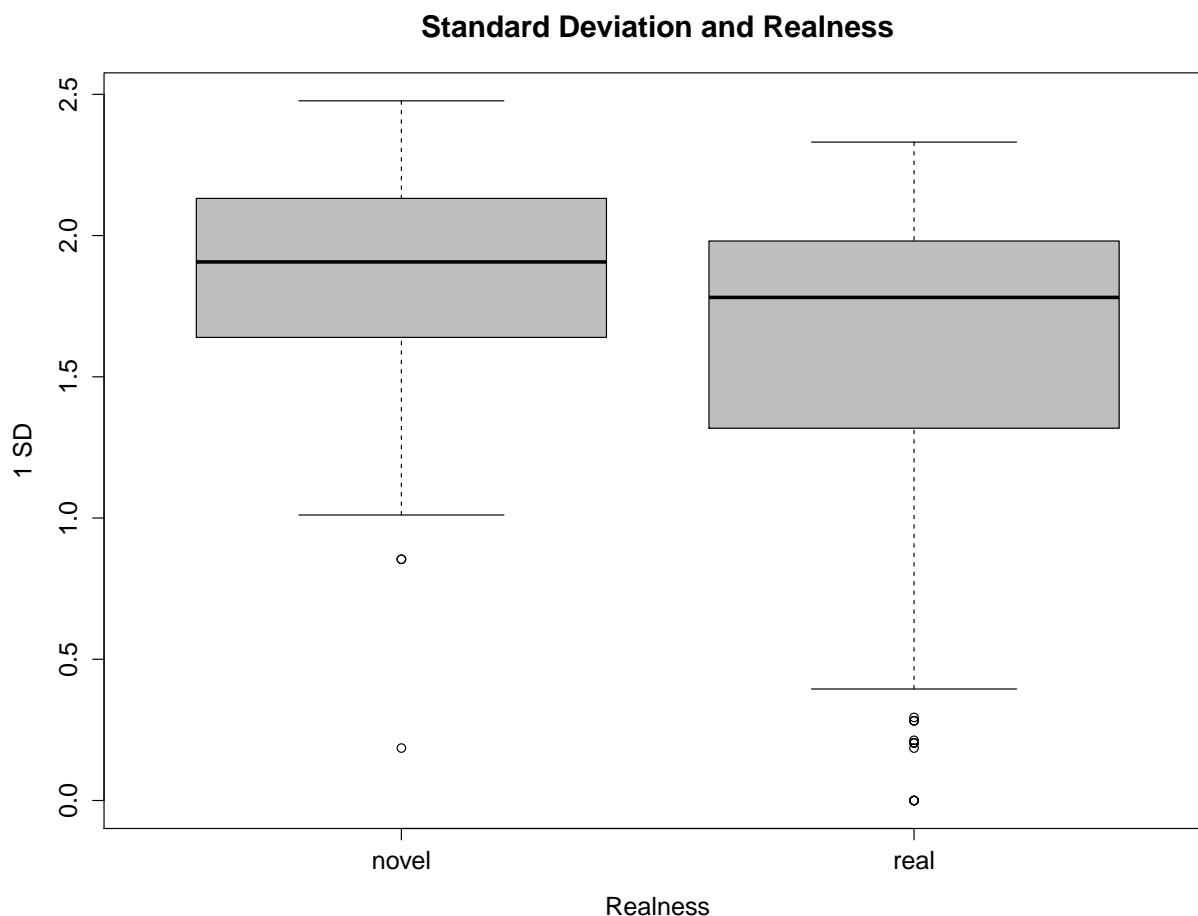


Figure 2.4: The standard deviation number of points (out of 7) according to realness (real English idiom vs. novel Scandinavian idiom).

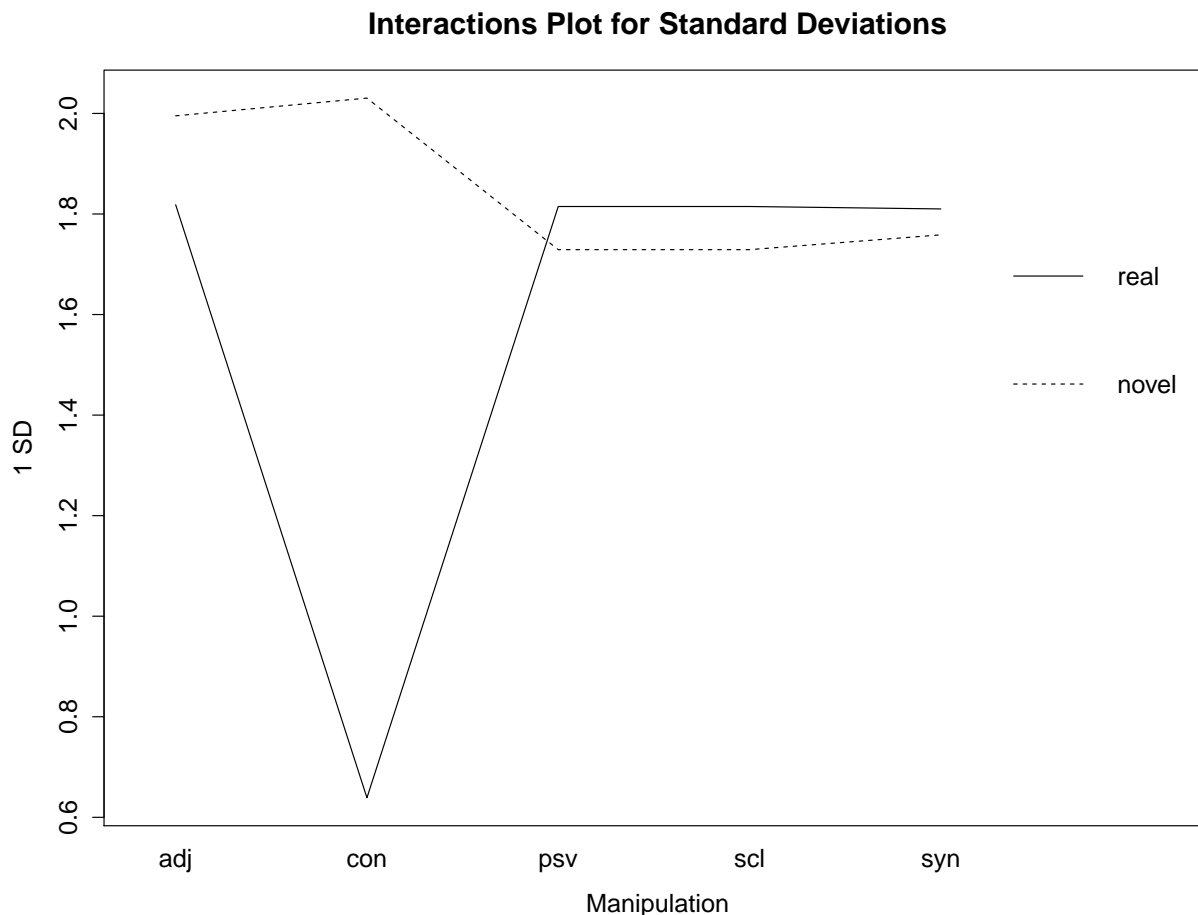


2.4.3 Judgment Consistency

In this section, I examine the consistency of participants' rankings for a given manipulation of a given idiom. This is the most pivotal data analysis of the experiment, since we determine from this whether competency with an idiom's flexibility requires idiosyncratic knowledge for each idiom. To do this, I calculated the between-participants standard deviation of the meaningfulness ratings of each manipulation of each idiom, producing the summary spreadsheet provided in Appendix B. The following analysis was performed on those item-wise standard deviations.

We can observe a significant main effect of realness, where the value of one standard deviation was higher for novel idioms ($M = 1.8480$) than for real idioms ($M = 1.579$, $F(1, 325) = 47.283$, $p \approx 0.000$). This finding, visualized in Figure 2.4, indicates that participants were in greater agreement regarding real idioms than novel idioms, *contra*. Tabossi et al. (2009b). The main effect of manipulation was also significant ($F(4, 325) = 24.207$, $p \approx 0.000$). *Post hoc* tests reveal that controls had a significantly lower standard deviation than adjective insertion, scalar substitution,

Figure 2.5: Interaction between realness (real English idiom vs. novel Scandinavian idiom) and manipulation for the value of 1 standard deviation. Key to conditions: *con*=control (“break the ice”); *adj*=adjective insertion (“break the stuffy ice”); *scl* = scalar substitution (“shatter the ice”); *syn* = non-scalar/synonym substitution (“split the ice”); *psv* = passivization (“the ice was broken”).



non-scalar substitution, and passivization, none of which were significantly different from each other. Finally, the two-way interaction between realness and manipulation was significant ($F(4, 325) = 27.185$, $p \approx 0.000$) and can be visualized on the interactions plot in Figure 2.5. *Post hoc* tests show that the real control averaged a significantly smaller standard deviation than all other conditions (all p values ≈ 0.000). Beyond this, the novel control averaged a significantly larger standard deviation than the novel scalar substitution condition ($p = 0.02$), the novel passivization condition ($p = 0.07$) and the novel non-scalar substitution condition ($p = 0.058$). The novel adjective insertion condition, whose average standard deviation does not differ significantly from that of novel controls ($p \approx 1.0$), has a significantly higher value than both novel scalar substitution ($p = 0.07$) and novel passivization ($p = 0.07$).

But how does this standard-deviation-based account of judgement reliability compare to the type of analysis that Tabossi et al. (2009b) performed? Tabossi et al. used binary acceptability judge-

ments (i.e. “Acceptable” or “Unacceptable,” with no intermediate values), then calculated what percentage of participants agreed with the majority. To see how my results compare to theirs, I performed a planned conversion of my 1-7 scale judgements to a binary scale, then performed the same analysis. Ratings from 1-3 were considered “unacceptable,” ratings from 5-7 were considered “acceptable,” and ratings of 4 were omitted from the analysis. As with Tabossi et al. (2009b), there was no significant main effect of realness ($F(1, 294) = 0.475, p = 0.491$). There was, however, a significant main effect of condition ($F(4, 291) = 20.98, p \approx 0.000$). Adjective insertion had the lowest average correctness, and this was significantly lower than scalar substitution ($p = 0.0046$), passivization ($p = 0.0064$), and the control condition ($p \approx 0.000$), but not for non-scalar substitution ($p = 0.811$). Non-scalar substitution, which had the next lowest average correctness, was significantly lower than the correctness for passivization ($p = 0.0763$) and the control condition ($p \approx 0.000$), and it approached a significant difference from scalar substitution ($p = 0.112$). Finally, the control condition showed greater correctness than both scalar substitution ($p \approx 0.000$) and passivization ($p < 0.0075$). Scalar substitution and passivization were not significantly different in terms of correctness ($p = 0.951$)

In addition to a main effect of manipulation, there was a significant two-way interaction between manipulation and realness in predicting correctness percentages ($F(4, 286) = 9.8877, p \approx 0.000$), visualized by the interactions plot in Figure 2.7. More subjects were “correct” for the real control than for any other condition (in all cases but novel-passivization, $p \approx 0.000$; for novel-passivization, $p = 0.0058$). In comparison to the least frequently “correct” manipulation, adjective insertion for novel idioms, three novel idiom manipulations were significantly more frequently correct: scalar substitution ($p = 0.00348$), control ($p = 0.00107$), and passivization ($p = 0.0179$). The difference in correctness between adjective insertion and non-scalar substitution for novel idioms was not significant ($p = 0.21$).

2.4.4 Judgment Patterns

The above analysis suggests that speakers are no more reliable at making acceptability judgments for idioms they know than for idioms they do not know. This leads us to believe that their syntax is not idiosyncratically stored on a per-idiom basis (see the discussion for a more in-depth analysis). Therefore, speakers must appeal to some principles in their competency with idiom flexibility. Here, we need to ask ourselves what patterns exist to constrain idiom flexibility.

Explanatory Factors

A linear probability model was used to determine which, if any, factors used to account for idiom flexibility show significant correlation with this judgment data. Since the real idiom stimuli were taken from Titone and Connine’s (1994) norms, measurements for the following properties were

Figure 2.6: Percentage of participants who agree with the majority, by manipulation. Key to conditions: *con*=control (“break the ice”); *adj*=adjective insertion (“break the stuffy ice”); *scl* = scalar substitution (“shatter the ice”); *syn* = non-scalar/synonym substitution (“split the ice”); *psv* = passivization (“the ice was broken”).

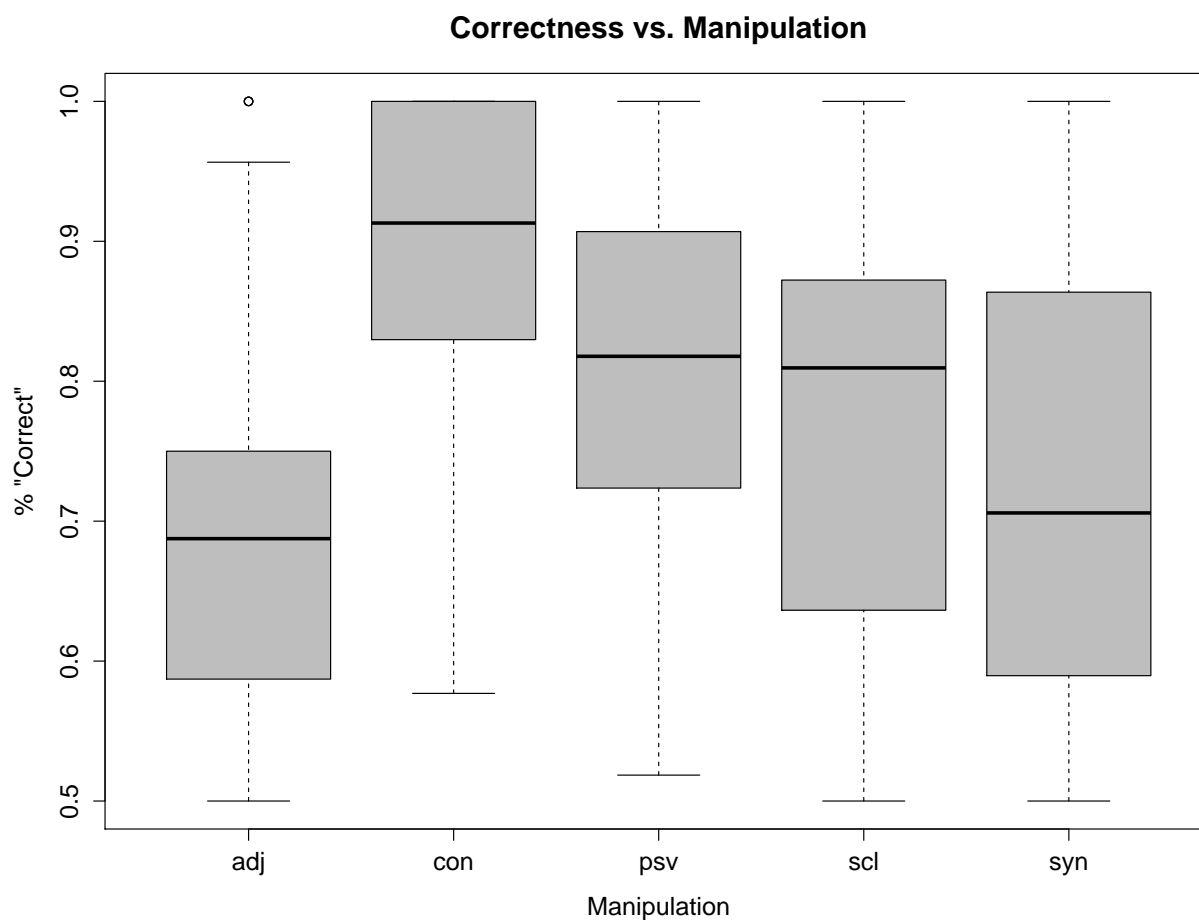
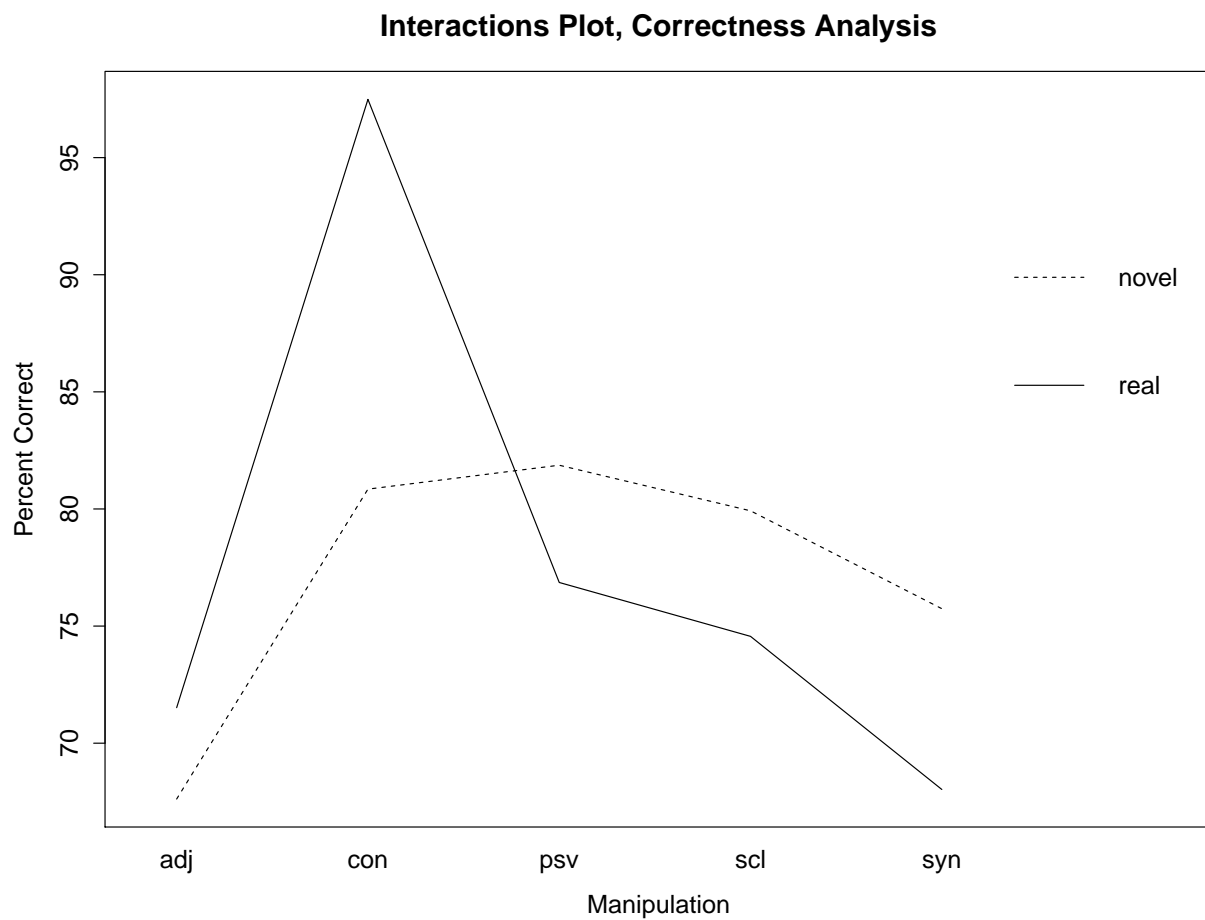


Figure 2.7: Interaction between realness and manipulation for the percentage of participants who agree with the majority. Key to conditions: *con*=control (“break the ice”); *adj*=adjective insertion (“break the stuffy ice”); *scl* = scalar substitution (“shatter the ice”); *syn* = non-scalar/synonym substitution (“split the ice”); *psv* = passivization (“the ice was broken”).



available for the real idioms: predictability, literality, meaningfulness, familiarity, and decomposability. *Predictability* refers to the rate in which participants are able to fill in the final word of the idiom from the previous words. *Literality* refers to the literal well-formedness of the idiom, or whether the idiom has a potential literal interpretation. For example, *kick the bucket* is literally well-formed, while *under the weather* is not. *Meaningfulness* refers to participants' self-assessment of how well they know an idiom. Finally, *decomposability* refers to the degree to which the individual component of the idiom contribute to the overall meaning.

In order to normalize and standardize the norming measures, z -scores were used in the linear model, where the control condition was treated as the base value. There was no main effect of decomposability, predictability, literality, meaningfulness, or frequency. Decomposability had a significant positive interaction with both adjective insertion ($p = 0.000267, \beta = 0.33630$)⁶ and non-scalar substitution ($p \approx 0.000, \beta = 0.41840$). Predictability showed a significant interaction for both passive ($p = 0.00863, \beta = -0.41628$) and scalar substitution ($p = 0.006601, \beta = 0.28117$). Literality showed a significant interaction with adjective insertion ($p = 0.006320, \beta = 0.2544$), scalar substitution ($p = 0.00436, \beta = 0.26518$), and non-scalar substitution ($p = 0.007778, \beta = 0.25116$). Meaningfulness interacted marginally with adjective insertion ($p = 0.061312, \beta = -0.30740$) and significantly with passivization ($p \approx 0.000, \beta = 3.17323$). Finally frequency interacted marginally with adjective insertion ($p = 0.90027, \beta = 0.29996$) and passivization ($p = 0.057879, \beta = -1.38412$).

In summary, we have seen that all of the factors under examination seem to have some correlation with idiom flexibility. However, they differ both in terms of which condition they were predictive for, and whether they were correlated with an increase or a decrease in flexibility for a given idiom. For the most part, the effect was only by fractions of a point (on the 1-7 scale), with the exception of meaningfulness and passivization. These results indicate that the factors under examination have only limited explanatory power; they explain at a level lower than the granularity of the Likert scale, only for some conditions. Also, these results suggest quite strongly that flexibility is not a singular property of an idiom. Rather, an idiom can be differently flexible for different conditions depending on different factors, some of which were examined here. While not a clean answer by itself, this tells us that we can expect a rather complex answer to the question of requisite knowledge for proper idiom use.

Novel Idiom Plausibility

For obvious reasons, the norming data we were able to use above to study real idioms simply does not exist for the novel idioms used in the experiment. However, for the novel idioms there exists a unique potentially interesting factor: *plausibility*. By using translations of true idioms in Finnish and Swedish, I intended for all of the novel idioms to be plausible. That is, they seem like phrases we could imagine entering the English language; they simply have failed to do so due to historical

⁶I use β to indicate the coefficient on the linear model, in order to illustrate its magnitude and direction.

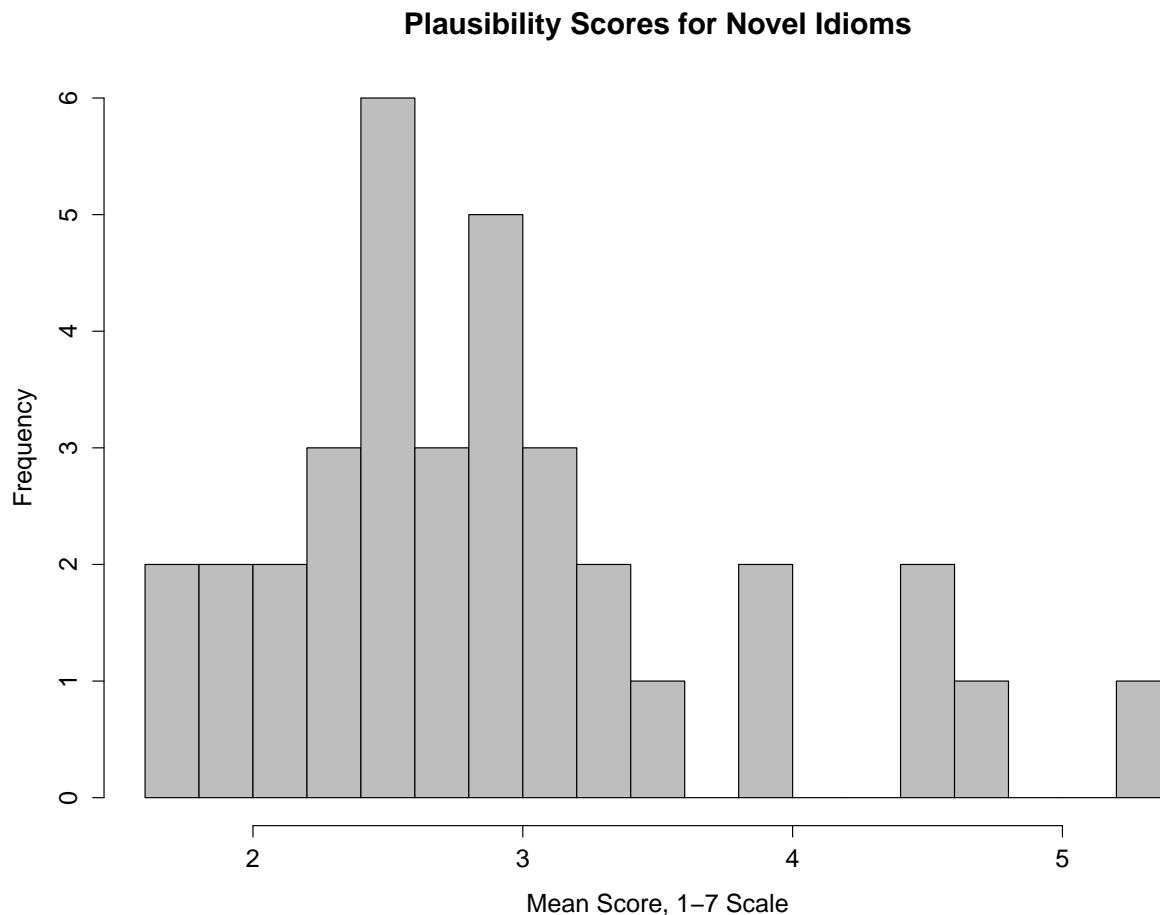


Figure 2.8: Distribution of the plausibility scores for the novel idioms used in the experiment.

reasons. Despite this, we can expect that speakers would not find them all equally plausible, and if speakers use principles to determine idiom flexibility, those principles might apply more reliably to more plausible idioms, facilitating greater flexibility.

To this end, I collected plausibility judgments on a 1-7 scale using Amazon Mechanical Turk. There were 44 respondents to the norming survey, 25 of whom provided data that fell within quality control standards. Each idiom's mean plausibility score was calculated, and the distribution of plausibility means can be seen in the histogram in Figure 2.8. Surprisingly, participants found most idioms to be relatively implausible, but perhaps this was result of participants preferring only idioms with transparent metaphorical meanings such as *snows of previous winters*. For a more detailed discussion, see Appendix C.

The mean plausibility for each novel idiom was included in a linear model where participants' acceptability response was treated as the dependent variable, and the independent variables were mean acceptability and condition. This model revealed a significant main effect of mean plausibility ($p = 0.0009, \beta = 0.26928$; in a between-subjects ANOVA, $F(1, 3800) = 136.1313, p \approx 0.000$).

Additionally, there were significant interactions between plausibility and adjective insertion ($p = 0.5699, \beta = 0.21880$), passivization ($p = 0.02368, \beta = 0.49951$), and non-scalar substitution ($p = 0.000204, \beta = 0.4278$). In general, then, we observe that more plausible idioms are more flexible, although not significantly so in the case of scalar substitution.

2.4.5 Examination of Items

Beyond providing a rich set of data for statistical analysis, the results of this survey include judgment data for 33 true idiomatic expressions in 3-4 different manipulations. Such material is perfect fodder for classic pencil-and-paper linguistic analysis, an arena where we can begin to look for patterns that merit further testing. Such patterns can lead to the uncovering of principles speakers use to determine idiom flexibility, so this data directly helps to answer the question at the core of this thesis. Rather than coming up with a hypothesis and seeking germane judgment-based data, we can move from the judgment-based data to hypotheses that more truly represent idiomaticity. This analysis will be a major component of Chapter 4, but I include some preliminary data here. I start with the sentences containing real idioms that were rated the most acceptable for each condition (24):

- (24) The most acceptable modified idioms
- a. **Adjective insertion:** Margaret is the real boss around here; she **calls the important shots**. (mean: 6.25, median: 7)
 - b. **Scalar substitution:** Gas is twice as expensive as it was ten years ago; now, it really **costs both arms and a leg!** (mean:5.9, median: 7)
 - c. **Non-scalar substitution:** Katie's memorization skills were impressive; there were fifteen violin concertos that she had **remembered by heart**. (mean: 6.29, median: 7)
 - d. **Passivization:** After Janice insulted him deeply, she **was given the cold shoulder** by Ricky for weeks. (mean: 6.83, median: 7)

We can see from these examples that in most of the cases (except perhaps (24a)), we might question whether the idiom is actually the full VP. For example, even though the Titone and Connine (1994) norms list *learn by heart* as an idiom, perhaps the only part that requires foreknowledge of the phrase is *by heart*; other verbs than *learn* that are compatible with some kind of internalization (e.g. *memorize, sing, recite*) are acceptable. Likewise, the actual idiomatic portion of *cost an arm and a leg* might just be *an arm and a leg*, and *the cold shoulder* is the idiom; it just happens to typically be *given* (probably in a light verb sense).

On the other hand, many of the idioms in the experiment were completely rejected in certain modified states, as exemplified by the lowest-rated real idioms of each condition given below:

- (25) The least acceptable modified idioms

- a. **Adjective insertion:** Stephen wasn't able to attend the meeting since he was feeling **under the viral weather**. (mean: 3.07, median: 2)
- b. **Scalar substitution:** Jesse was very focused on searching for a job, but she couldn't help but keep her relationship's impending failure **at the middle of her mind**. (mean: 1.64, median: 1)
- c. **Non-scalar substitution:** Patrick knew many high-profile band managers, so it didn't surprise us when he got front-row tickets to the sold out concert **out of light air**. (mean: 1.93, median: 1)
- d. **Passivization:** Erica meant to stop by the post office, but **her mind was slipped by it**. (mean: 1.5, median: 1)

The reasons that a modification might not succeed are many, which renders generalizations over examples such as those in (25) virtually impossible and possibly not worthwhile. In the case of (25c), it is not clear that *light* is a good synonym for *thin* in *out of thin air*; its failure might just be a lack of appropriate synonyms. In the case of (25d), the complete lack of agency to the “subject” *it* probably causes the inability to passivize, and (25b) probably does not establish adequate context for a scalar substitution to be understood. It might be the case that *under the weather* (25a) is truly just unable to meaningfully accept adjectives; this example might be more informative for pencil-and-paper analysis.

We have only been able to glimpse at the wealth of judgment data available from this experiment, data that will prove essential to understanding the various factors influencing idiom decomposability and idiom flexibility in general. I will return to a close examination of this judgment data in Chapter 4, discussing particularly some of the factors explaining the great flexibility of some idioms, as well as some of the reasons why many other idioms failed to accept certain modifications. With more than 100 example sentences, even this will only scratch the surface, but the use of pre-existing judged sentences instead of sentences constructed to illustrate a point should offer us methodological comfort in this endeavor.

2.5 Discussion

Due to the scope of this experiment, it brings us a wealth of results. I will first discuss the implications of speakers' agreement on judgments for a given idiom, then the implications of the judgments themselves, and finally what the results say about the source of speaker's judgments of idiom flexibility.

We want to know whether a necessary component of using an idiom properly is knowing its idiosyncratic lexical syntactic flexibility. In this experiment, this was measured by examining participants' agreement on the flexibility of idioms with which they were familiar (“real idioms”) and “idioms” with which they were not familiar (“novel idioms”). Looking simply at main effects of realness (i.e.

whether an idiom is a true English idiom or a Scandinavian calque), we might be initially tempted to say that participants were more consistent in their judgments for real idioms than for novel idioms. When we look at standard deviations on a 7-point scale instead of percentage of agreement on a binary scale, the greater resolution of our data seems to produce this result. However, a closer examination shows that this effect is driven entirely by the control condition for the real idioms (Figure 2.5). Unsurprisingly, participants are more consistent in saying that a sentence containing the canonical form of an idiom (e.g. *He broke the ice*) is acceptable than any modified form of the idiom. It is somewhat surprising that participants were no more reliable for the novel controls than for the other novel conditions, but perhaps this reflects a difficulty in accepting novel idioms generally or concerns about their plausibility.

Regardless, the point of this analysis was to see if participants differ in how reliably they make acceptability judgments for novel idioms in comparison to real idioms. The best conclusion seems to be that participants are equally reliable at judging real and novel idioms, which suggests that syntactic flexibility is not idiosyncratically specified in the lexicon. While we need to be careful in drawing conclusions from a null result, the fact that the real controls performed differently from the rest of conditions lends some credence to the idea that participants are equally reliable in their acceptability judgments regardless of foreknowledge of the idiom. **Therefore, on the whole, this study has confirmed Tabossi et al.’s (2009b) results for a more diverse set of idioms with a more fine-tuned judgment scale.**

Now we turn to the judgments themselves. Most importantly, I predicted that participants would generally find scalar substitution (e.g. “The old man nudged the bucket”) to be acceptable. We might specify this hypothesis by saying that the mean acceptability rating for scalar substituted idioms would exceed 4, the middle of the Likert scale. This hypothesis was contradicted by the experiment. Indeed, scalar substitution was the least acceptable of all of the conditions, with a mean rating of 2.783. However, as I suggested in my brief discussion of the individual items in Section 2.4.5, the issue might not be that scalar substitution is not acceptable. Rather, the stimuli did not provide a sufficient context to enable scalar substitution. Of course, one could make this argument about all failed modifications, but my intuition is that scalar substitution often requires a very particular context: a previous statement of the canonical form of the idiom within the discourse. We saw this already for the *kick the bucket/nudge the bucket* alternation, replicated in (26a). If the discourse did not contain a non-modified form of *kick the bucket*, the substitution becomes far more marginal (26b). We can observe a similar alternation for a decomposable idiom with the NP substituted (27).

- (26) a. Tom: *Did the old man kick the bucket last night?*
 Joe: *Nah, he barely nudged it* (McGlone et al., 1994)
- b. ?#The old man barely nudged the bucket last night.
- (27) a. He didn’t just let the cat out of the bag; he let out **the whole litter!**

- b. ?#He let **the whole litter** out of the bag!

Syntactic flexibility does not seem to be lexically pre-specified, and the jury remains out on scalar substitution. However, these two initial questions only constitute a fraction of the understanding that this data set brings. Looking at the explanatory factors of decomposability, literal well-formedness, predictability, meaningfulness, and familiarity, we found that these factors do offer some explanatory value. However, they only show a significant effect for some manipulations. For example, decomposability was a significant predictor for both adjective insertion and non-scalar substitution, but not for scalar substitution or passivization. Predictability showed the opposite trend, where it was significant for passivization and scalar substitution, but not adjective insertion or non-scalar substitution. This double dissociation reveals that idiom flexibility is not only based on several factors; these factors interact differently depending on the operation. This result poses problems to some existing formal views that place idioms into a hierarchy of levels of flexibility, since it seems clear from this result that no idiom has a single value for “flexibility”; it might accept modification A but not modification B, while a different idiom might accept modification B but not modification A.

The coefficients for the linear model also should give us pause. Except for passivization’s interactions with meaningfulness and frequency, the effect of the metrics such as decomposability is significant but minor in magnitude. This is important, since a strong form of Gibbs and Nayak’s (1989) Idiom Decomposability Hypothesis, and certainly Jackendoff’s (1997) model of the lexicon, predict that decomposability alone would be a very strong predictor of flexibility in all manipulations. Beyond this, while most of the time the coefficients were positive, the coefficients for passivization and predictability and for meaningfulness and adjective insertion were negative. In other words, more predictable idioms are *less* likely to accept passivization, and more meaningful idioms (meanings are best known by subjects) are less likely to accept adjective insertion.

In short, idiom flexibility seems to function as an interaction of multiple principles with varying pull depending on the manipulation and context. There is considerable variability in speakers’ judgements for idiom flexibility, but this variability itself does not correspond to any interesting patterns. Particularly, whether a speaker has ever heard the idiom before does not seem to matter for the reliability of judgments, even if they tend to judge all new idioms as less flexible. Although this suggests that some principles must exist, the lack of context in the scalar substitution sentences prevented us from establishing any particular principles.

2.6 Conclusion

On the whole, this study confirmed Tabossi et al.’s (2009b) results showing that participants have equally reliable judgments for the flexibility of real and novel idioms. This study was able to confirm this for a far less idealized set of idioms with much finer-grained judgment data, so we

can now be confident that idiom syntactic flexibility does not result from idiosyncratic stored properties of idioms. Returning to the original question of what speakers need to know in order to competently use and understand idioms, we have learned that foreknowledge of the individual flexibility information about each idiom is *not* part of this requisite knowledge, so speakers must know certain principles in order to form opinions regarding idiom flexibility.

This finding does not directly tell us anything about what principles exist. Indeed, our attempt at illustrating a principle for some idiom flexibility, that of scalar substitution, was not confirmed by the data. However, this was likely due to the lack of proper context in the stimuli. Regardless, a regression analysis involving factors including decomposability showed that some perceived properties of given idioms can be considered minutely predictive of their flexibility for certain *but not all* modifications. In summary, then, we can abandon two potential proposals for idiom flexibility: (1) that it is idiosyncratically stored and (2) that it is a singular property of an idiom. An idiom might be “flexible” on one dimension (like adjective insertion) but not on another (such as word substitution). A speaker cannot possibly know “how flexible an idiom is” as part of their linguistic competence, since no such measure can truly exist. Rather, they use a variety of principles to determine whether *a given modification* is acceptable.

The large set of judgment data allows us to move forward with postulating some principles of the multi-dimensional concept of idiom flexibility, and such analysis can be found in Chapter 4. These principles matter crucially in these thesis since they represent a component of a speaker’s competence with idiom flexibility. However, before we continue, it is important to ask why decomposability has been viewed as so central to all questions of idiom flexibility. Experiment 2 examines this most popular postulated principle of what speakers use to determine idiom flexibility: breaking parts of the idioms meaning across the words in the phrase. Can speakers even do this reliably? In order to resolve this lingering question, a question of why a seminal study (Gibbs and Nayak, 1989) would be at odds with my results, Experiment 2 was performed (Chapter 3).

Chapter 3

Experiment 2: Skeletons in the Closet

3.1 Background

The Idiom Decomposability Hypothesis (Gibbs and Nayak, 1989) has long been the primary principle behind idiom flexibility in the literature. At its core, the proposal suggests that to know how to use an idiom – the central question of this thesis – one needs to know how to decompose idioms (or whether a given idiom is decomposable). But do speakers actually do this reliably? We have seen from both our preliminary examination of the Idiom Decomposability Hypothesis in the introduction and from a rigorous examination of a diverse set of idioms in the first experiment that the Idiom Decomposability Hypothesis fails to account for much of the diversity found in idiom flexibility. Moreover, it has received very limited corroboration in subsequent research beyond the proposals of Jackendoff (1995, 1997, 2002), as discussed in Tabossi et al. (2009a).

Despite these challenges, the concept of idiom decomposability provides an intuitive principle by which speakers analyze idioms, so qualifying the original hypothesis in light of our data should prove to be a useful path forward. Before we consider qualifying the Idiom Decomposability Hypothesis, we should understand why it did succeed in the context of Gibbs and Nayak’s (1989) experiments. I have already alluded to part of a reason: Gibbs and Nayak use deliberately idealized items. They went through a list of idioms compiled in a previous study and selected the idioms that they thought were most certainly normally decomposable, abnormally decomposable, and nondecomposable. Thus, in their Experiment 1, where they find that subjects do reliably categorize idioms according to decomposability, they were merely corroborating their own judgments.

However, a critical examination of their stimuli suggests that even within this idealized design there exists a worrisome confound. Gibbs and Nayak presented each idiom with an approximate meaning, and the correspondence between the syntactic argument structure of the meaning and the argument structure of the idiom was confounded with decomposability. For example, most of the decomposable idioms of the syntactic type $[_{VP} V NP]$ were presented with meanings that were

also of the syntactic type [VP V NP]. As we can see from Gibbs and Nayak’s (1989) decomposable idiom stimuli replicated in Table 3.1, *rack one’s brains* is paired with ‘to search one’s memory,’ and it is easy to see from a purely syntactic perspective that *rack* corresponds to *search* and *one’s brains* corresponds to *one’s memory*. Of the twelve “decomposable” idioms, there are only two exceptions to this trend: *button one’s lip* (‘to stop talking’) and *pat on the back* (‘to give praise’). When there is a nearly one-to-one correspondence between the syntactic argument structure of an idiom and its meaning, it is much easier for an individual to imagine, especially in the context of an experimental task, that semantic decomposability holds.

Idiom	Meaning
pop the question	to propose marriage
lay down the law	to give strict orders
break the ice	to start a conversation
let off steam	to release tension
rack one’s brains	to search one’s memory
lose one’s grip	to lose control
miss the boat	to lose opportunity
hit the sauce	to drink liquor
play with fire	to experiment with danger
button one’s lip	to stop talking
clear the air	to resolve dispute
pat on the back	to give praise

Table 3.1: Normally decomposable idioms and their meanings, as presented to participants in Gibbs and Nayak’s (1989) Experiment 1.

Gibbs and Nayak’s participants also reliably considered the idioms presented in Table 3.2 to be nondecomposable. The lack of correspondence between the syntactic argument structure of the idiom and its meaning in Table 3.2 stands in stark contrast to the “decomposable” idioms in Table 3.1. Only two idioms come close to corresponding syntactically with their meanings: *play the field* (‘to date several people’) and *knock on wood* (‘to hope for luck’). The rest, such as the pairing of *kick the bucket* with ‘to die,’ have no such correspondence. Thus, when a participant is faced with the question of whether these idioms are decomposable, s/he might be disinclined to say so, because the meaning presented inhibits the mapping. Notably, we can see a way in which some of these idioms might approach decomposability. For example, *pack a punch* might be conceptualized as ‘to possess power’ rather than ‘to be powerful,’ in which case *pack* would correspond to the meaning ‘possess’ and *punch* would correspond to the meaning ‘power.’ Moreover, *give the sack* and *give the bounce* are probably best construed as light verb idioms,¹ where the idiomatic meanings are housed within the words *sack* and *bounce* (this will be discussed in greater detail in the next chapter). Thus, we have reason to question whether all of the “nondecomposable” items truly cannot be construed as having parts of the idiom bear parts of the meaning.

¹See discussion in Chapter 4.

Idiom	Meaning
kick the bucket	to die
chew the fat	to talk aimlessly
cool one's heels	to wait impatiently
make the scene	to arrive at party
go for broke	to risk everything
give the sack	to fire from job
play the field	to date several people
shoot the breeze	to talk casually
pack a punch	to be powerful
knock on wood	to hope for luck
give the bounce	to get rid of

Table 3.2: Nondecomposable idioms and their meanings, as presented to participants in Gibbs and Nayak's (1989) Experiment 1.

I hypothesized that Gibbs's results were based on the correspondence between the syntactic argument structures of the idiom and the given meaning, rather than inherent decomposability of the idioms. The experiments below test this hypothesis by presenting the same idioms as Gibbs used but with varied meanings. In the remainder of this chapter, I present two experiments, which I will refer to as Experiments 2A and 2B, respectively. They are both designed to test this hypothesis, differing only in the choice of filler stimuli and secondary task. With some qualification, both versions of the experiment fail to support my hypothesis, showing that Gibbs and Nayak's (1989) results for their first experiment obtain even when controlling for the confound of syntactic matchedness. This means that speaker judgments of idiom decomposability are real, not task-induced, which tells us that some version of decomposability can be appealed to in understudying the principles speakers use in determining idiom flexibility.

3.2 Experiment 2A

The goal of this experiment was to test to what extent participants' ratings of decomposability are determined by the correspondence between the syntactic argument structure of the idiom and the syntactic argument structure of the meaning presented alongside it (henceforth, "matchedness"). If so, judgments of decomposability were merely task effects in Gibbs and Nayak's (1989) original study. To accomplish this, participants were presented with a series of idioms taken from the Appendix of Gibbs and Nayak's (1989) study, along with meanings that were either syntactically matched or syntactically unmatched to the idiom (see Section 3.2.1). They were then asked to determine whether the idiom was decomposable or nondecomposable.

A challenge to discussing this experiment is that there are two values that we might call "decomposability." First is the category that Gibbs and Nayak (1989) assigned to a given idiom (an independent variable) and second is the designation that each participant gave an idiom (the de-

pendent variable). For the sake of clarity, I will refer to the former as the “decomposability label” and the latter as the “decomposability response” or “decomposability rating.”

3.2.1 Methods

Participants

A total of 77 participants were recruited through Amazon Mechanical Turk. Originally, we targeted 100 participants in order to permit statistically significant tests after eliminating subjects with unreliable data (see Section 3.2.2). However, preliminary analysis suggested that some significant effects were already visible, so we stopped seeking further participants by canceling the HIT in the interest of running a modified version (Experiment 2B). Only MTurk workers who have completed at least 1000 HITs with a 95% or higher approval rating were granted access to the task. They were paid \$0.25 for an average of 4 minutes and 49 seconds worth of work. Participants’ HITs were approved regardless of whether their data was deemed acceptable, though they were not told that this would be the case.

Materials

Ten of Gibbs and Nayak’s (1989) decomposable idioms and ten of their nondecomposable idioms were paired with both argument-matching and non-argument-matching meanings, a variable that I refer to as “matchedness.” By “argument matching,” I refer to the idea that the syntactic argument structure of the idiom (e.g. [VP V NP]) corresponds to the syntactic argument structure of the given meaning. To the extent that was possible, one of these meanings was copied directly from the Gibbs and Nayak stimuli. To facilitate this, argument matching was interpreted loosely such that idioms of the form [VP V NP] or [VP V PP] simply needed a meaning with the same form, allowing for multi-word verbs like *lay down* in *lay down the law* and the use of adjectives within NPs. In other words, the internal structures of the NPs and PPs in the meanings of the “matched” condition did not necessarily align exactly with the internal structures of the idiom itself, but they at least shared the form [VP V NP]. The meaning pairs for Gibbs and Nayak’s (1989) list of “decomposable” idioms can be found in Table 3.3, and the meaning pairs for the nondecomposable idioms are provided in Table 3.4.

Additionally, novel idiom fillers were selected to be used in a quality control task. Ten of the VP novel “idiom”-meaning pairs used in Experiment 1 were arbitrarily selected for these purposes. From all of these materials, two experimental lists were devised. In each list, the 10 idioms that Gibbs and Nayak (1989) deemed normally decomposable and for the 10 idioms that they considered nondecomposable, 5 were presented with argument-matched meaning, and 5 were presented with an argument-mismatched meaning. Both lists contained all 10 novel “idioms.”

Table 3.3: Normally decomposable idioms used in the experiment, with their syntactically matched and syntactically unmatched meanings. Definitions taken from Gibbs and Nayak’s (1989) Experiment 1 are marked with an asterisk.

Idiom	Matched	Unmatched
pop the question	to propose marriage*	to propose
lay down the law	to give strict orders*	to be authoritative
break the ice	to start a conversation*	to be congenial
rack one’s brains	to search one’s memory*	to think hard
lose one’s grip	to lose control*	to become incapable
miss the boat	to lose opportunity*	to be unsuccessful
hit the sauce	to drink liquor*	to drink excessively
play with fire	to experiment with danger*	to live dangerously
button one’s lip	to shut one’s mouth	to stop talking*
clear the air	to resolve dispute*	to clear up

Table 3.4: Nondecomposable idioms used in the experiment, with their syntactically matched and syntactically unmatched meanings. Definitions taken from Gibbs and Nayak’s (1989) Experiment 1 are marked with an asterisk.

Idiom	Matched	Unmatched
kick the bucket	to end life	to die*
chew the fat	to discuss random things	to talk aimlessly*
make the scene	to attend a party	to be social
give the sack	to carry out a job termination	to fire from job*
play the field	to date several people*	to date extensively
shoot the breeze	to discuss random things	to talk casually*
pack a punch	to possess power	to be powerful*
knock on wood	to hope for luck*	to be hopeful
give the bounce	to cause departure	to get rid of*
raise the roof	to display anger	to get angry*

Procedure

After selecting the HIT on Amazon Mechanical Turk, subjects were directed to a link to a survey made through and hosted by the Qualtrics survey software. They were presented with a welcome screen explaining the length and demand of the task, followed by a second screen that more thoroughly explained their task. The instructions that they were given are below (boldface in original). These instructions were adapted from Titone and Connine’s (1994) instructions for decomposability in the creation of their norms. They are consistent with Gibbs and Nayak’s description of their instructions and use example idioms that were not presented in the experiment.

In this survey, you will be presented with a series of “idioms” and their approximate meaning. Some of these idioms will be real; others will not.

For the “**idioms**” that you don’t recognize, you will select “Not a real idiom.”

For **the idioms that you do recognize**, you will determine whether it is decomposable. An idiom is “decomposable” if the individual components of the idiom made some unique contribution to the phrases figurative paraphrase. An idiom is “nondecomposable” if its individual components do not make such a contribution.

An example of a **decomposable idiom** is *save your skin*, which approximately means “to protect yourself.” This idiom is considered decomposable because the word *save* corresponds to the part of the meaning “protect,” and *your skin* corresponds to the part of the meaning “yourself.” If you are presented with an idiom like this, select “decomposable.”

An example of a **nondecomposable idiom** is *cool your heels*, which approximately means “to wait impatiently.” This idiom is nondecomposable because the individual words (*cool*, *your*, *heels*) do not directly relate to the overall meaning of the idiom. If you are presented with an idiom like this, select “nondecomposable.”

Participants were presented with one of the two lists of all 20 idioms. Additionally, each participant was presented with the 10 novel-idiom pairs. These idioms were presented in random order for each subject using the Qualtrics survey software. For each idiom, they selected one of three choices: “Decomposable,” “Nondecomposable,” and “Not a real idiom.” At the end of the survey, participants were given the opportunity to leave comments. They were then given a confirmation code to copy into the MTurk HIT interface.

3.2.2 Results

The results of this experiment were analyzed using the R statistical programming language (R Development Core Team, 2011). First, the data from any unreliable participants were excluded (Section 3.2.2). After this, a linear probability model was used to test for the effect of Gibbs and Nayak’s (1989) designations of decomposability and my designation of syntactic matchedness on participant’s judgments of decomposability (Section 3.2.2). Finally, the idioms themselves were examined using a linear probability model (Section 3.2.2). In general, no effect was found for syntactic matchedness of the presented meaning, and very few individual idioms showed any effect at all, suggesting that speakers’ judgments of decomposability are not task effects of the presented meaning.

Quality Control

Due to the nature of crowdsourced research, several quality control measures, reminiscent of Experiment 1, were taken. I used the following metrics: survey completion time, instructional reading time, successful identification of non-real idioms, and successful identification of real idioms. Participants’ data were excluded if they completed the survey more than 1.5 standard deviations above or

below the mean completion time. Obviously, any participant who completed the study too quickly could have been randomly clicking, and participants who took too long might not have been focusing on the task (after all, they are using their web browser in any number of locations). Participants' data was also eliminated if the amount of time they took to read the instructions for the task was more than 1.5 standard deviations above or below the mean. We cannot trust their responses if they did not read the directions, but if they took too long to read the directions, this might be a sign that they did not understand the task (after all, the experimenter was not present to answer questions).

Beyond these timing-based quality assurance metrics, I used the task itself to ensure data quality. Participants were expected to mark the novel idioms in the task as “Not a real idiom” (rather than “decomposable” or “nondecomposable”); if they failed to do this more than once, their data was eliminated from further analysis. After participants were removed based on timing and false positive identifications of idioms, I removed data where they incorrectly identified real idioms as “Not a real idiom.” However, there is a risk that some of the idioms used by Gibbs and Nayak (1989) are no longer common and would therefore not be recognized by participants. Therefore, I first removed from analysis all idioms that more than 25% of the remaining participants identified as fake. This led to the removal of *give the sack*, *give the bounce*, and *make the scene*, all idioms that Gibbs and Nayak had labelled “nondecomposable.” With these unrecognized idioms eliminated, any participant who incorrectly identified more than two real idioms as “not real” was excluded from further analysis. This number was selected to allow for the fact that even fully-engaged participants who are correctly completing the task might not recognize every idiom.

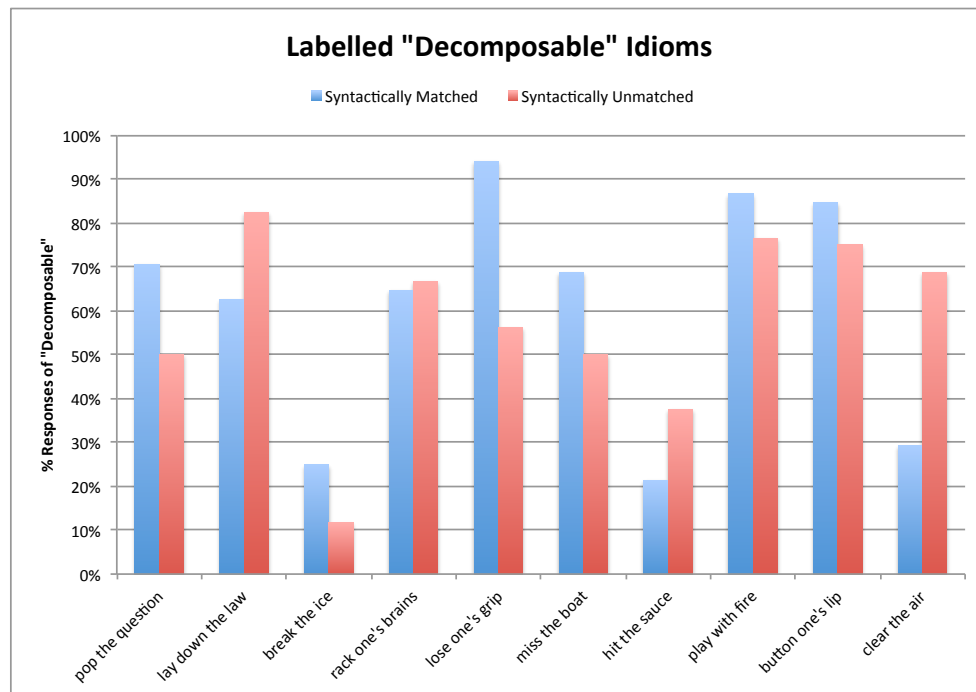
In the end, the data from 33 participants remained for use in the subsequent analysis.

Decomposability and Matchedness

The remaining data was studied using a linear probability model. A main effect of decomposability label was observed ($p \approx 0.000$), with an average of 35% more participants providing a decomposability response of “decomposable” for idioms that were labeled “decomposable” than for idioms that were labelled “nondecomposable.” There was no significant main effect of syntactic matchedness (the correspondence between the syntactic structure of the idiom and its presented meaning), and no significant interaction between matchedness and decomposability label.

This means that my hypothesis (that matchedness explains decomposability responses) was not upheld by this study. Indeed, even the labelled decomposable idioms in the unmatched condition were given a positive decomposability response by significantly more participants than the labelled nondecomposable idioms in the matched condition ($p \approx 0.000$).

Figure 3.1: Idioms that Gibbs and Nayak (1989) labelled “Decomposable,” with percentage of participants in Experiment 2A who considered them decomposable.



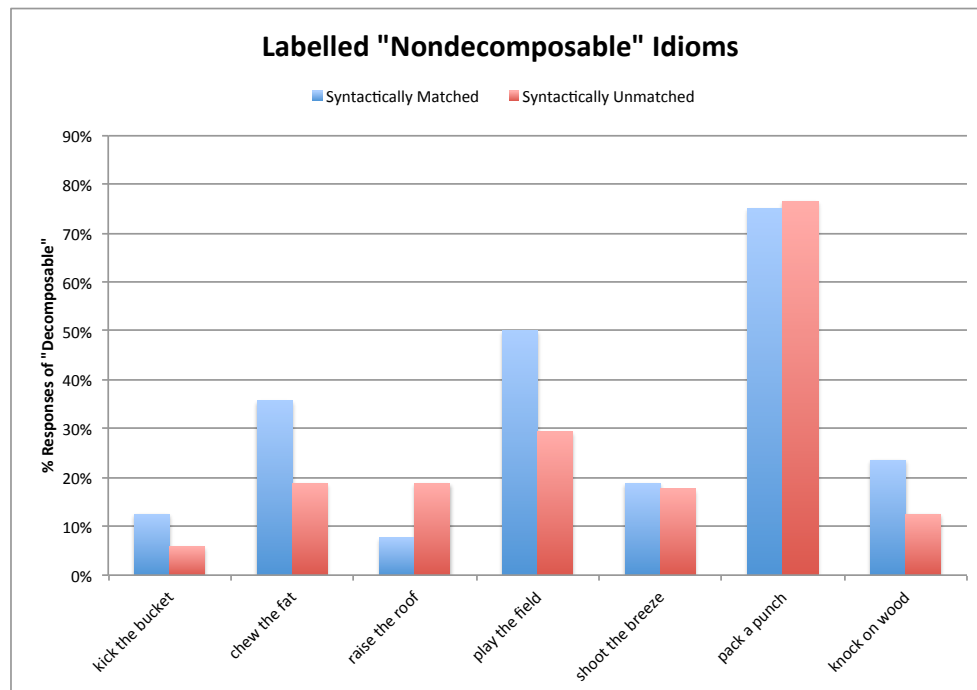
Items

Figures 3.1 and 3.2 show the percentage of decomposability responses by idiom. Some of the idioms in both of these graphs show some difference between the matched and unmatched condition, and the significance of these differences was examined using a linear probability model. This model treated decomposability response as the dependent variable, and examined the interaction between syntactic matchedness (true or false) and idiom, where each idiom was treated as a factor. A significant interaction was observed only for *clear the air* ($p < 0.015$), where the syntactically *unmatched* condition was more frequently considered decomposable by participants. (There is still no main effect of matchedness when we exclude *clear the air* from analysis.)

It appears that the majority of participants disagreed with Gibbs and Nayak’s (1989) decomposability labels for at least three idioms. The labelled decomposable idioms *break the ice* and *hit the sauce* were considered decomposable by less than half of participants in both syntactic matchedness conditions, and *pack a punch*, which Gibbs and Nayak labelled as nondecomposable, was considered decomposable by over 75% of my participants for each syntactic matchedness condition.

Finally, simple visual inspection of Figures 3.1 and 3.2 reveals that participants often disagreed with each other about whether an idiom was decomposable. Gibbs and Nayak (1989) found that 86% of participants responded that the labelled decomposable idioms were decomposable. Here, we find this for only 59% of participants (68% when *break the ice* and *hit the sauce* are excluded).

Figure 3.2: Idioms that Gibbs and Nayak (1989) labelled “Nondecomposable,” with percentage of participants in Experiment 2A who considered them decomposable.



Likewise Gibbs and Nayak found that 88% of participants agreed on the nondecomposability of the idioms they had labelled “nondecomposable,” while the present study found that this held for only 71% of the labelled “nondecomposable” idioms (79% if we exclude *pack a punch*). Thus, we find that participants did not agree on decomposability response to the same extent that participants in Gibbs and Nayak’s original study did.

3.2.3 Discussion

In general, Gibbs and Nayak’s (1989) results were confirmed in this study. Participants generally agree that the idioms Gibbs and Nayak labelled as decomposable are, in fact, decomposable, and the idioms they label as nondecomposable are, in fact, nondecomposable. Participants have these judgments regardless of how the meaning is stated, since no effect of syntactic argument matchedness is observed. Indeed, syntactic matchedness produced an effect for only one of the idioms (*clear the air*), and in this case, the syntactically matched condition was given considerably fewer decomposability responses than the syntactically unmatched condition. This result probably does not stem from matchedness, but rather from the fact that the syntactically unmatched meaning (‘to clear up’) contains a word from the idiom, while the syntactically matched meaning (‘to resolve dispute’) does not. Indeed, the only labelled decomposable idiom that visually (but not significantly) favors the syntactically matched condition is *lose one’s grip*, which contained the word *lose* in the syntactically matched but not the syntactically unmatched meaning.

Despite the confirmation that Gibbs and Nayak’s influential results were not the result of a confound, the present results offer some qualification to the original ones. We see that some idioms have apparently waned in popularity such that many participants no longer recognized them. More importantly, participants’ judgments regarding decomposability clearly differ from those in the original study.² Not only were three idioms apparently miscategorized by Gibbs and Nayak (1989); participant agreement regarding decomposability was rarely overwhelmingly consistently in the direction of decomposability or non decomposability. Rather, decomposability was a moderately weak trend.

Although the general lack of matchedness effects suggests that the confound I observed in Gibbs and Nayak’s study did not impact the results, there is a risk that the lack of observed matchedness here effects is actually a task effect. Since participants needed to rely on their pre-existing knowledge of the idioms to perform the quality-control task of identifying non-real idioms, they had no reason to actually read the given meaning of an idiom.³ Indeed, they might have perceived this as a waste of time. Experiment 2B addresses this concern.

3.3 Experiment 2B

The results of Experiment 2A do not seem to align with my hypothesis that syntactic argument correspondence determines speakers’ judgments of decomposability. However, the task could have led them to focus on their own knowledge of the idiom, rather than paying attention to the meaning given in the experiment. Experiment 2B was adapted from Experiment 2A in order to mitigate this concern.

Specifically, the quality-checking filler task for Experiment 2B was altered, preserving the critical stimuli from Experiment 2A. Half of the idiom-meaning pairs contained “mismatched” meanings, and participants were required to identify such pairs as such, sorting all other idioms they were familiar with according to decomposability. Since reading the meaning was necessary to perform the quality control task in a satisfactory manner, we can be confident that any participants who perform acceptably on the filler quality control task have read the provided meanings.

3.3.1 Method

Participants

A total of 90 participants were recruited through Amazon Mechanical Turk, with the same requirements as Experiment 2A. MTurk workers were told not to take the HIT if they had participated in

²Gibbs and Nayak (1989) do mention that three idioms in their study did not exhibit the level of agreement of the rest, but they never identify the idioms.

³The effect for *clear the air* might indicate otherwise, but perhaps the visual of having a word in both the idiom and the meaning would have an effect even if participants were not focusing on the given meanings.

the previous study. Those who had participated in Experiment 2A (based on their MTurk WorkerIDs) were excluded from analysis. Since participants were presented with twice as much filler for this task as they were for Experiment 2A, the payment per HIT was increased to \$0.30 for an average of 6-7 minutes of work.

Materials

The experimental stimuli (including lists) were identical to those of Experiment 2B. However, the filler stimuli were changed. Exactly twenty real VP idioms of English that did not appear in the critical stimuli were paired with meanings taken from the novel idioms of my Experiment 1. They were deliberately paired with meanings that were not even tangentially related to that of the actual idiom, such as the pairing of *hit the sack* and ‘to speak up.’ This helped to ensure that participants would not be overly particular about the subtle meaning differences that necessarily resulted from the construction of the critical stimuli. In other words, by ensuring that the meaning-mismatched filler pairs were severely mismatched, I minimized the risk that participants would select “Meaning is totally wrong” for loose meanings like ‘to be congenial’ for *break the ice*. The use of the phrases “**approximate** meaning” and “**totally** wrong” (emphasis added) also helped to mitigate this concern.

3.3.2 Procedure

As with Experiment 2A, each participant was presented with one of the two lists of twenty critical stimuli. Additionally, each participant saw all twenty mismatched real idiom pairs. Since all presented idioms were real (i.e. correct identification of real idioms was not an important matter of quality control), participants were given the option to decline to categorize any idiom with which they were completely unfamiliar. They were also given explicit instructions for the task of identifying mismatched idiom/meaning pairs, replicated below (boldface in original):

In this survey, you will be presented with a series of idioms and an approximate meaning.

If you are **completely unfamiliar with the idiom**, select “I do not know this idiom.” Please be honest; this is not a component of HIT approval.

In some cases, the approximate meaning is totally incorrect. If the “approximate meaning” is nowhere close to the actual meaning of the idiom, select “Meaning is totally wrong.” If the meaning is pretty close but not exact, follow the following directions for approximately correct pairings.

For **the idioms paired with an approximately correct meaning**, you will determine whether they are decomposable. ...

The above lines preceded the explanation of decomposability, as shown for Experiment 2A. For each idiom, participants were presented with the idiom, a meaning, and four choices: “Decomposable,” “Nondecomposable,” “Meaning is totally wrong,” and “I do not know this idiom.” All idiom-meaning pairs, both critical and noncritical, were presented in random order.

3.3.3 Results

With the exception of adapted quality control based on the new mismatch approach, identical analyses were run on the results of Experiment 2B as they were for 2A, again using R (R Development Core Team, 2011). Here, I will summarize the changes in quality control (Section 3.3.3), analyze labelled decomposability and matchedness (Section 3.3.3), and examine items.

Quality Control

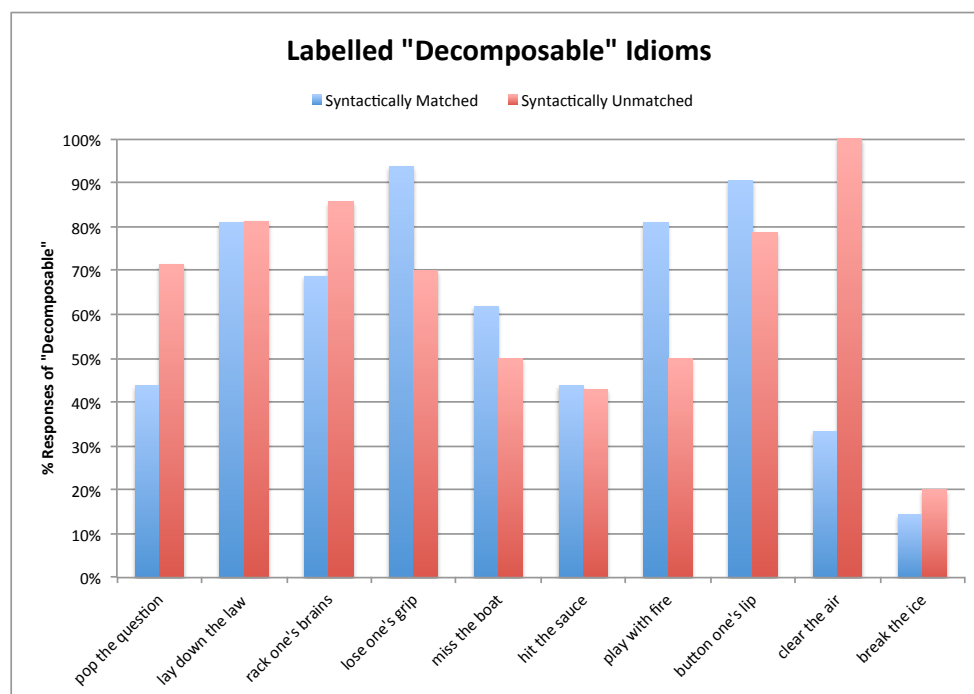
Participant exclusions based on survey completion time and instruction reading time were identical to those used for Experiment 2A. The procedures for exclusion based on false positives and false negatives were adapted to the new experimental task. Participants were excluded if they identified more than one of the twenty completely incorrect meanings as correct. Following this, data were excluded from idioms that for more than 25% of the time were either not recognized or mistakenly considered to be paired with a correct meaning. This eliminated the following idioms: *raise the roof*, *give the bounce*, and *make the scene*, all idioms that Gibbs and Nayak categorized as nondecomposable.

Following this, participants were excluded if they did not recognize more than two idioms and/or if they identified more than one correct meaning as incorrect. After all of these exclusions were performed, 40 participants remained for further analysis. Of these, 3 had participated in Experiment 2A, and their data were subsequently excluded, leaving 37 participants.

Decomposability and Matchedness

The remaining data was studied using a linear probability model. A main effect of decomposability label was observed ($p \approx 0.000$), with an average of 38% more providing a decomposability response of “decomposable” for idioms that were labeled “decomposable” than for idioms that were labelled “nondecomposable.” There was no significant main effect of syntactic matchedness, and no significant interaction between matchedness and decomposability label. This means that the results of Experiment 2A are not, in fact, task effects. Rather, whether or not the syntax of the phrase and its meaning correspond has no effect. Indeed, just as in the case of Experiment 2A, even the labelled decomposable idioms in the unmatched condition were given a positive decomposability response by significantly more participants than the labelled nondecomposable idioms in

Figure 3.3: Idioms that Gibbs and Nayak (1989) labelled “Decomposable,” with percentage of participants in Experiment 2B who considered them decomposable.



the matched condition ($p \approx 0.000$).

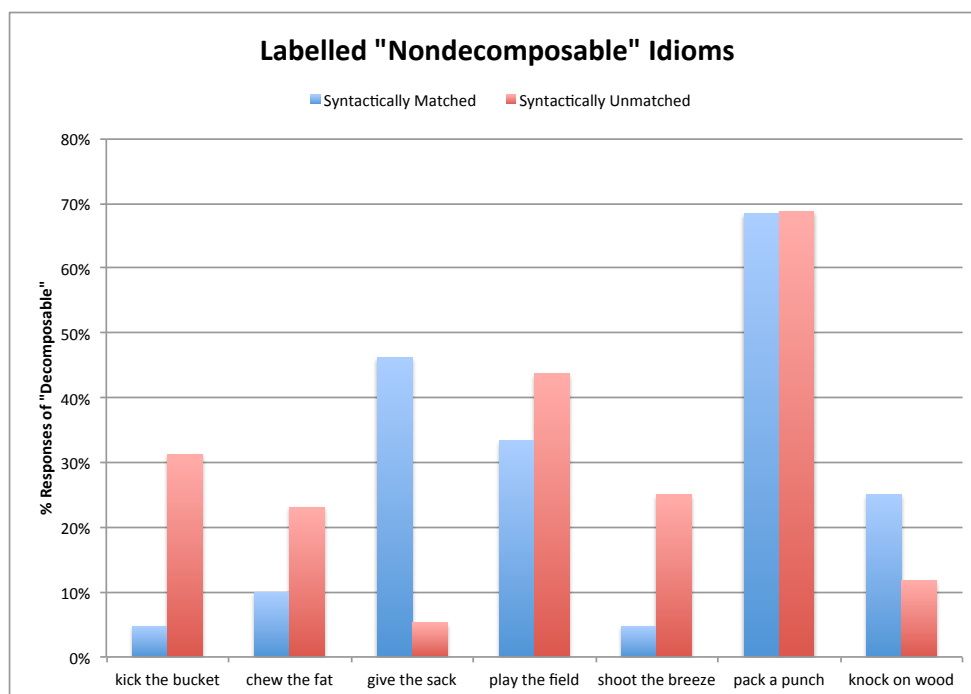
Items

Figures 3.3 and 3.4 show the percentage of decomposability responses by idiom. Some of the idioms in both of these graphs show some difference between the matched and unmatched condition, and the significance of these differences was examined using a linear probability model. This model treated decomposability response as the dependent variable and examined the interaction between syntactic matchedness (true or false) and idiom, where each idiom was treated as a factor.

Significant interactions were observed for *clear the air* ($p < 0.00453$) and *play with fire* ($p < 0.085$), which are both labelled “decomposable,” as well as *give the sack* ($p < 0.035$), which is labelled “nondecomposable.” As in Experiment 2A, the effect for *clear the air* was the reverse of what had been expected; the syntactically unmatched condition received more ratings of decomposable than the unmatched condition, presumably for the same reasons (word correspondence between the conditions). The other two idioms, however, exhibited the expected trend; the syntactically matched condition was more frequently rated decomposable than the syntactically unmatched condition.

Again, some idioms seemed to be miscategorized. Both *hit the sauce* and *break the ice* were rated as decomposable by fewer than 50% of participants, despite being labelled “decomposable” by Gibbs and Nayak (1989). On the other hand, *pack a punch* was considered decomposable by nearly 70%

Figure 3.4: Idioms that Gibbs and Nayak (1989) labelled “Nondecomposable,” with percentage of participants in Experiment 2B who considered them decomposable.



of participants, despite being labelled “nondecomposable.” Beyond this, participant agreement on which idioms were decomposable varied considerably. Only 63% agreed that labelled decomposable idioms were decomposable (cf. 86% for Gibbs and Nayak (1989)), and 71% agreed that labelled nondecomposable idioms were nondecomposable (cf. 88% for Gibbs and Nayak (1989)). If we remove the apparently miscategorized idioms, these numbers increase to 71% for labelled decomposable idioms, and 78% for labelled nondecomposable idioms.

3.3.4 Discussion

The results of this experiment generally recapitulated those of Experiment 2A. Most importantly, no general effect of matchedness was observed, even when we are confident that participants have read each idiom’s meaning. However, we did see that in some cases, the wording of the meaning affected the idiom’s decomposability rating. The higher decomposability rating of *clear the air* in the unmatched condition again probably came from the use of the word *clear* in both the idiom and the unmatched meaning. However, the other two other idioms which showed an effect of matchedness probably do not face the same issue. The word *with* does overlap in *play with fire* and its syntactically matched meaning (‘to experiment with danger’), but if decomposability holds true, *with* preserves its literal meaning in the context of the idiom. Regardless, *give the sack* was rated more decomposable in the matched condition, even though the overlap was far less definite (‘to carry out a job termination’ vs. ‘to fire from job’). This result might have to do with the fact that

give the sack is a light verb idiom, which I will discuss briefly in the General Discussion (Section 3.5) and in greater depth in the next chapter.

Beyond this, this study confirmed that *pack a punch*, *break the ice*, and *hit the sauce* were miscategorized by Gibbs and Nayak, at least according to my participants. Additionally, *give the bounce* and *make the scene* were not recognized by at least 25% of participants in both versions of Experiment 2. The only new idiom not to be recognized was *raise the roof*. This is probably because participants thought that its meaning ('to display anger'/'to get angry') was mismatched, since the idiom has recently taken on a new meaning associated with partying.

3.4 Pooled Analysis

Since both versions of Experiment 2 had essentially the same result, it was possible to pool the data and run the same analyses as above using all 70 reliable participants. Again, decomposability label was significant in the linear probability model ($p \approx 0.000$), but syntactic matchedness of the meaning was not ($p = 0.763$). In terms of individual idioms, *clear the air* was rated decomposable significantly more in the syntactically unmatched condition than in the syntactically matched condition ($p \approx 0.000$), and there was a marginally significant effect for *lose one's grip*, which was rated decomposable significantly more in the syntactically matched condition than in the syntactically unmatched condition ($p = 0.0844$).

3.5 General Discussion

Overall, both versions of Experiment 2 failed to show any significant effect of syntactic matchedness between idioms and their stated meanings. On the other hand, Gibbs and Nayak's (1989) designations of decomposability did show a significant main effect. These experiments confirm that a confound in the original study did not actually bias the results. When we look past the general effects to the specific items, however, the present studies do offer some qualification of the original study. First, some of the items are no longer well known, perhaps exhibiting an accelerated pace of linguistic change when it comes to idioms; around 25 years has passed between the studies. Second, participants systematically disagreed with Gibbs and Nayak's characterization of *pack a punch* as nondecomposable; they seem to think that *pack* corresponds to 'possess' and *punch* corresponds to 'power' – even if the meaning they are presented with is 'to be powerful.'

Participants also had the impression that *break the ice* and *hit the sauce* were nondecomposable, a result that I find more surprising. In both cases, the issue might be that clear English words do not exist to capture parts of the meaning. For example, *break the ice* was paired with 'to start a conversation' or 'to be congenial,' but it more closely means 'break down a (fragile rigid) barrier to social interaction' (Jackendoff, 1997). The mapping of *ice* to the meaning 'a (fragile rigid) barrier to

social interaction’ probably creates extra difficulty. Additionally, *hit the sauce* was paired with the meanings ‘to drink liquor’ and ‘to drink excessively.’ If we want to decompose the idiom, however, *the sauce* refers to the drinks, while *hit* is functioning something like a light verb. Perhaps we should paraphrase ‘to have alcoholic beverages.’

These explanations for the (non)decomposability ratings of *break the ice* and *hit the sauce* point to a potentially important finding of both of these studies: at times, the way we state the meaning of an idiom can affect speaker’s decomposability judgments. While this is purely speculative for *break the ice* and *hit the sauce*, the significant interaction between type of definition and the idioms *clear the air*, *play with fire*, and *give the sack* clearly illustrates the potential for the statement of the meaning to influence speakers’ judgments of decomposability. This influence, however, does not seem to be a pure effect of syntactic matchedness. Rather, it involves other factors like exact word correspondence. If my explanation for *break the ice* and *hit the sauce* is correct, this might mean that a “matched” meaning needs not only be syntactically matched, but semantically decompose in the same way. Thus, *give the sack* seems decomposable when paired with ‘carry out a job termination,’ since *give* corresponds to ‘carry out’ and *the sack* corresponds to the ‘job termination.’ However, *break the ice* does not seem decomposable when paired with ‘to start a conversation,’ since *break* does not correspond semantically to ‘start,’ and *ice* does not correspond semantically to ‘conversation.’

But how should we interpret this intermittent effect? On one hand, we could see this as evidence that our conceptualizations of idioms are actually somewhat flexible, sometimes able to be molded by suggestion. The strongest evidence for this is *give the sack* in Experiment 2B. On the other hand, we might argue that the general lack of an effect of syntactic matchedness proves that speakers have a rigid conceptualization of each idiom, regardless of the approximate definition they see. Under this view, the occasional effect of definition phrasing did not reflect lexical storage, but rather the nature of the task, which required higher level linguistic analysis – analysis that we might roughly label as “meta-linguistic.” The wording of the instructions might have accentuated these effects, since the meaning of the example decomposable idiom was in both syntactic and semantic correspondence with the idiom’s internal structure. These kinds of questions cannot necessarily be avoided in judgment-based tasks, but they are critical to our understanding. I will explore this point in greater detail in the next chapter.

These issues all point toward the question of how idioms are conceptualized– are their conceptualizations flexible, and how do their conceptualizations interact with semantic structure and syntactic structure? This could be a partial answer to the question of what speakers need to know in order to properly use an idiom; particularly, they would need only an imprecise notion of its meaning that could be contextually adapted. The field is nowhere near the point of providing complete answers, but I will use the next chapter to further explicate the issue and point toward future directions of possible research.

3.6 Conclusions

We have seen that speakers usually judge the decomposability of an idiom reliably regardless of how its meaning is stated to them, but with some exception by idiom. But what has this experiment told us about what speakers need to know in order to properly use an idiom?

First, it has shown us that the word “properly” might be problematic, considering the amount of disagreement that exists regarding given idioms’ categories. This study raises a question about speaker variation in idiom understanding. Clearly, there was considerable disagreement among participants regarding the decomposability of idioms, and the simple binary response did not capture an individual participant’s relative judgments of decomposability. For example, we might ask whether subjects were less confident of their categorization of idioms whose decomposability response was nearer to 50% than for those idioms whose decomposability response was closer to 0% or 100%.

Along these lines, we have qualified our understanding of decomposability. It is possible that part of knowing how to use idioms properly is knowing how to decompose them, but the level of disagreement that we have noted would suggest that decomposability can only play a partial role.

Finally, the general lack of effect of presented meaning (with exceptions for overlapping words) suggests that speakers have reliable conceptualizations of an idioms’ meaning and how the syntactic and semantic structures link. That is, to know an idiom means to have a clear conceptualization of its meaning. This suggests that idioms probably are not constructed on-line.

Knowing now that idiom syntactic flexibility is not stored idiosyncratically and that decomposability offers an incomplete picture, in the final chapter of analysis, I turn to some general principles that seem to be involved in idiom flexibility and representation, the essential components of knowing how to use an idiom.

Chapter 4

Where the Rubber Meets the Road

In this final chapter of analysis, I look at the trends we can observe from the items in Experiments 1 and 2. In particular, I focus on trends in the representation and flexibility of idioms that the data in Experiment 1 allow us to make. These principles and their lexical implications are at the core of what it means to understand how to use an idiom. I will start by discussing patterns I feel are explained by understanding some idioms to be idioms embedded with collocations, making a sketch of a proposal for their lexical representation (Section 4.1). After this, I briefly discuss the role of conceptual metaphors in idiom flexibility (Section 4.2), continuing with a sketch of what these mean for the lexicon (Section 4.3). In short, this chapter addresses both how idioms are represented in the lexicon and (some of) the source of their flexibility: fundamental aspects of what speakers need to know in order to use idioms.

4.1 Idioms and Collocations

4.1.1 Collocations containing Idioms

The study of idomaticity is plagued with issues of boundaries: boundaries of what constitutes as decomposable, what constitutes as an idiom at all, etc. Another, less appreciated issue, is the boundaries of the idioms themselves. As we saw in Experiment 1, some of the most flexible idioms were *cost an arm and a leg*, *learn by heart*, and *give the cold shoulder*. All of these idioms were taken from Titone and Connine's (1994) norms, so there is clear linguistic literature support for the existence of these as idioms. However, I proposed that the boundaries are incorrect; the idioms – the parts of the phrases that must be stored separately in order to be understood – actually are sub-parts of these phrases. For example, instead of *cost an arm and a leg*, the idiom is actually *an arm and a leg*, which roughly means ‘a very expensive price.’ Since ‘expensive prices’ almost always are *cost*, the phrase usually appears as *cost an arm and a leg*. But this is just a collocation.

This issue extends beyond the handful of idioms mentioned in the analysis of Experiment 1; looking through Titone and Connine's (1994) list of 171 idioms, at least twenty-four of these idioms strike me as actually being a collocation containing an idiom. In most cases, the collocation's verb is grouped as part of the idiom, while the actual idiom is either an NP¹ (28) or a PP (30). Light verb idioms constitute a particularly interesting subset of VP collocations containing NP idioms (29), and they will be addressed in greater detail in the next subsection. In all of the examples below, I place brackets around the portion of the collocation that I take to be the actual idiom.

- (28) NP (Titone and Connine, 1994)
- a. be [a wet blanket], be [the cat's whiskers], be [the spitting image]
 - b. keep [a level head], keep [an ace up your sleeve]
 - c. back to [square one]
 - d. climb on [the bandwagon]
 - e. cost [an arm and a leg]
 - f. pay [lip service]
- (29) Light verb(-ish) constructions (Titone and Connine, 1994)
- a. get [the eye], get [the picture], get [your goat]
 - b. give [plenty of rope], give [the cold shoulder], give [the creeps], give [the sack], give [the willies]
 - c. have [a fling], have [an axe to grind], have [cold feet]
- (30) PP (Titone and Connine, 1994)
- a. go [against the grain]
 - b. learn [by heart]
 - c. sit [on the fence]
 - d. skate [on thin ice]

From the above examples, it seems clear that many conventionalized phrases exist where only a part needs to be understood idiomatically in order for the phrase to be meaningful. This seems to be one clue toward understanding the nature of some multi-word expressions, both in terms of their flexibility and in terms of their mental representations. To understand this clue in greater depth, I turn briefly to a discussion of the subclass of light verb idioms.

4.1.2 Light Verb Idioms

Idiomatic constructions containing a light verb prove to be an important subset of these constructions. I have alluded to the existence of "light verb idioms" throughout this thesis, since they seem

¹I have placed the brackets around the DP rather than the NP, but I am agnostic about whether the "actual idiom" is the DP or the NP, or whether the question matters at all.

to exist in the nebulous space of semi-productivity. At the most basic level, light verb constructions (LVCs) are multi-word expressions, generally of the form [_{VP} V NP], in which the verb has become semantically bleached, serving a mostly grammatical function to permit the complement NP to function like a verb (Jespersen, 1965). Some common English light verbs include *give*, *take*, *make*, *do*, and *have*, as illustrated by the examples below:

- (31) a. He took a walk.
 b. He walked.
- (32) a. Everyone but Joey took a ride in the hot air balloon. (North, 2005)
 b. Everyone but Joey rode in the hot air balloon.

The verb does make some semantic contribution, however, since *He gave a tour* does not mean the same thing as *He toured* (North, 2005). The degree to which such expressions “need” to exist in the lexicon remains a matter of open debate, since different NPs do seem to require different light verbs with only partially clean patterns. Computational modeling has provided some clues (North, 2005), but the psycholinguistic question remains difficult, and I will only begin to address it here.²

In the present studies, a handful of light verb idioms have entered the picture. In Experiment 1, *give someone the cold shoulder* almost certainly constituted a light verb idiom, and one could certainly argue that *get the picture* and *hit the sack* resemble LVCs, even if they are not exactly the same. More exactly, they probably exist on the least semantically bleached end of the light verb continuum, which Kearns (2002) refers to as “Vague Action Verbs.” Within the Gibbs and Nayak (1989) stimuli used for Experiment 2, there exist several potential LVCs, all of which Gibbs and Nayak had identified as “nondecomposable”: *make the scene*, *give the sack*, and *give the bounce*. Unfortunately, these expressions were not well-known to participants, so *make the scene* and *give the bounce* were excluded in both versions of the experiment, and *give the sack* was analyzed only for Experiment 2B.

We should observe that *give the sack* is indeed a fairly clean example of an LVC. The word *sack* as a verb means ‘to fire’ (perhaps only in British English), so *give the sack* shows the same pattern as the most prototypical light verb examples given in (31) and (32):

- (33) a. My boss finally gave the incompetent IT guy the sack.
 b. My boss finally sacked the incompetent IT guy.

Despite the relative dearth of light verb idiom data for Experiment 2, *give the sack* proved to be an incredibly interesting item, as it was one of the two idioms for which my anticipated pattern of syntactic argument matching proved significant. When the given meaning was ‘to fire from job,’ very few (about 5%) participants found the idiom to be decomposable, but when the meaning was

²See Butt (2003) for an in-depth overview of the complexities of light verbs.

stated as ‘to carry out a job termination,’ the number increased drastically (to 46%). Why was this the case?

First, it could be challenging even to a linguist to determine whether *give the sack* should be considered decomposable. Instead of having the parts of the meaning correspond to parts of the phrase, the entire idiomatic meaning is condensed into the word *sack*; *give* makes a very limited semantic contribution. If the meaning is ‘to sack,’ then the semantic contribution of *sack* is ‘sack,’ while *give* makes no semantic contribution. Thus, only one part of the idiom seems semantically decomposable, an option which typical definitions of idiom decomposability fail to permit.

This sort of confusion probably underlies the effect observed in Experiment 2B. When participants see the meaning ‘to fire from job,’ they think about the definition of decomposability provided in the task and decide that *give the sack* does not fit. However, when the definition is stated as ‘carry out a job termination,’ the correspondence becomes transparent, and ‘carry out’ is semantically vacuous enough to correspond to *give*. In other words, the meaning ‘to fire from job’ blocks a decomposable understanding, while ‘to carry out a job termination’ facilitates this interpretation, even if only for a task of explicit decision (vs. sentence acceptability).

I introduced this subsection by noting that light verb idioms constitute a (proper) subset of collocations containing idioms. By this I mean that the actual idiom in a light verb idiom is the NP, such as *sack*. However, unlike the case of *cost an arm and a leg*, the whole phrase needs to be present in order for the idiomatic meaning to be recognized. For example, both *give the sack* and *hit the sack* are idioms, but the word *sack* means vastly different things in each case. In the former, it refers to a ‘job termination,’ while in the latter, it seems to refer to a ‘place of sleep.’ Therefore, we need to know the entire phrases, even if purely on the level of ambiguity resolution. In the case of *give the sack*, there also exists the complement *get the sack*, which roughly means ‘to be fired’ or perhaps more precisely ‘to receive a job termination.’ This expression obviously serves as a direct foil to *get the sack*, showing the very limited semantic contribution a verb can make in an LVC (34a). It seems that this conceptualization of *sack* is constrained to those two light verbs, however; if we substitute the non-light *receive* for the light verb *got*, the idiomatic meaning of the construction becomes unavailable (34b).

- (34) a. The incompetent IT guy finally **got the sack**.
 b. # The incompetent IT guy finally **received the sack**.

4.1.3 Implications for Flexibility

I have shown that many expressions frequently referred to as “idioms” actually contain parts that are idiomatic and parts that are merely collocative in nature. Thus far, I have relied primarily on definitions of idioms, decomposability and light verb expressions, as well as my own intuitions about the CI’s meanings. However, these points could all be moot if the categories do not lead to

different patterns in the linguistic and psychological data. The results from Experiment 1 provide such linguistic data. The following CIs were used in the study: *give the cold shoulder*, *cost an arm and a leg*, *learn by heart*, *get the picture*, and *hit the sack*.

If only a portion of a CI is truly “idiomatic,” than we could expect that CI flexibility would be very high when modifications occur collocation-internally but not idiom-internally, but much less high idiom-internally, all other things being equal. On the whole, this prediction plays out in the data. Most of the sentences containing CIs and receiving the best judgments included collocation- but not idiom-internal modification. Remarkably, many of these sentences received median acceptability scores of 7, meaning that they were just as acceptable as the citation form of the CI:³

- (35) ⁷After Janice insulted him deeply, she **was given the cold shoulder** by Ricky for weeks.
- (36) ^{5.5} By the time we finished explaining the conflict, Evan **got the ugly picture**.
- (37) ⁷Katie’s memorization skills were impressive; there were fifteen violin concertos that she had **remembered by heart**.
- (38) ⁷Katie’s memorization skills were impressive; there were fifteen violin concertos that had been **learned by heart by her**.
- (39) ⁷Gas is twice as expensive as it was ten years ago; now, it really **sells for an arm and a leg!**

This was not universally true, however. Generally, it seems that *hit the sack* was less flexible (40). Perhaps it is closer to a typical idiom than a CI, although separate reasons could explain these patterns. For example, the scalar substitution example (40a) tried to create a contrast by placing a light verb on a scale, which perhaps is unreasonable for a semantically impoverished word (41a). Passivization (40b) might just be semantically/pragmatically odd even if the idiomatic meaning is available. Adjective insertion was neutral (40c), but its lack of complete acceptability could be explained if the idiom is not only the NP *sack* but the DP *the sack*. This leaves us with the unavailability of passivization for *cost an arm and a leg* (41a), another case where passivization simply makes no sense (41b).

- (40) a. ¹Hannah’s day at work was very tiring, but since she had evening plans, she could only **touch the sack** when she arrived at home.
- b. ²Hannah’s day at work was so tiring that **the sack was hit by her** as soon as she arrived at home.
- c. ⁴Hannah’s day at work was so tiring that she **hit the cozy sack** as soon as she arrived at home.
- (41) a. ³Gas is twice as expensive as it was ten years ago; now, **an arm and a leg is cost by it!**

³Here, I adopt the convention of putting the median 1-7 acceptability scale rating as a superscript preceding the sentence in lieu of the traditional */?/[good] system used in linguistic literature.

- b. ?*Lots of money **is cost by gas** nowadays!

On the other hand, almost all of the CI sentences containing an idiom-internal modification were given acceptability ratings on the bottom half of the scale, as exemplified below. Word substitutions of the scalar variety (42) and non-scalar variety (43) were among the least acceptable examples, and the one case where adjective insertion occurred NP-internally was also generally considered unacceptable (44).

- (42) a. ^{2,5}We explained the conflict so thoroughly that by the time we were done, Evan **got the whole gallery**.
 b. ²Katie's had a remarkable ability to internalize music; there were fifteen violin concertos that she had **learned by soul**.
- (43) ¹Hannah's day at work was so tiring that she **hit the bag** as soon as she arrived at home.
- (44) ^{3,5}After Janice insulted him deeply, Ricky **gave her the cold offended shoulder** for weeks.

Such inflexibility has exceptions. For example, *an arm and a leg* seems to be somewhat flexible internally (45). In the supposedly scalar case of (45b), the modification is nearly external, and the insertion of the adjective "pricey" (45a) does not add much to the meaning; it simply reinforces the meaning of 'something very valuable.'

- (45) a. ⁶Gas is twice as expensive as it was ten years ago; now, it really **costs an arm and a pricey leg!**)
 b. ⁷Gas is twice as expensive as it was ten years ago; now, it really **costs both arms and a leg!**

In summary, the CIs generally accept modification readily when this modification is collocation-internal but not idiom-internal. When the modification is idiom-internal, acceptability judgments drastically decrease. This result supports the idea that certain phrases that have been labelled as "idioms" actually constitute a more complex type of multi-word expression.

But what does this fact tell us more broadly about idioms and the lexicon? Perhaps here is a case where Jackendoff's (1997) representation of decomposable idioms could prove useful if we allow for the idea that collocations, like idioms are lexical items (Jackendoff, 1995). Below is an example for *give the sack*, modeled on Jackendoff's (1997) analysis of *bury the hatchet*:

- (46) Lexical item for *give the sack*
 a. **Morphology and Phonology:** (_aWd /gɪv/) (_bCl /ðə/) (_cWd /sæk/)
 b. **Syntax:** _aV_x NP_y

$$\begin{array}{c} \wedge \\ \textit{bD} \quad \textit{cN} \end{array}$$

c. **Semantics:** $[GIVE_{LV}([\]_A, [JOB\text{-}TERMINATION]_y)]_x$

This raises an obvious question of the connection in the lexicon between *the sack* in *give the sack* and *the sack* in *get the sack*. I contend that this is exactly the same question on the relationship between *bucket* in *kick the bucket* and *bucket* alone, which has been addressed variously (Jackendoff, 2002; Tabossi et al., 2009b) in the literature. Unfortunately, this thesis is only able to focus on idioms lexical representation as it stands, rather than as it relates to other lexical items. However, this is certainly part of a speaker’s idiom competency, and it could provide further answers as to what speakers need to know about an idiom in order to properly use it.

4.2 Conceptual Metaphors, Conceptual Structure

The above section describing collocation-idioms explains a fraction of what speakers know in order to properly use a subset of multiword expressions. The principles and representations proposed move us toward a greater understanding of both the memorized material and principles associated with idiom/collocation usage. However, these principles do not scale directly to other idioms. The present section takes a step back, looking at broader principles of idioms’ interaction with the conceptual (and even visual) system to explain what kinds of principles must come into play when we manipulate idioms. More than anything, it is to the broad principles presented in this section that speakers probably appeal when determining an idiom’s flexibility.

4.2.1 Idiom Analyzability

In Experiment 1, we looked at several factors that have some predictive value in different types of idiom flexibility: predictability, literality, meaningfulness, familiarity, and decomposability. Additionally, we saw that the plausibility of novel idioms positively influenced their acceptability in modified form. Many of these factors partially capture a broader generalization regarding speakers’ understanding of idioms: *according to a variety of dimensions, speakers find some idioms to be easier to internally analyze than others*. Decomposability captures the fact that speakers can connect parts of the form to parts of the meaning. Meaningfulness approximates speakers’ general understanding of what an idiom means, which might be related to the connection they can draw between the literal words and the way the phrase is used. Literal well-formedness permits internal analysis on the level that it allows a clear image to be evoked, and facilitates the existence of a normal compositional interpretation. Predictability and familiarity less directly relate to speakers’ understandings of the source of an idiom’s meaning, but perhaps they simply facilitate the identification of an idiom, even in modified form.

The brilliance of Nunberg’s (1978) proposal of decomposability lay in the fact that he realized that differences in idioms exist at the interface level. In his proposal, decomposability indicated

a correspondence within an idiom at the syntax-semantics interface, where we can think of “semantics” in the formal linguistic sense of the term. However, a third component of the cognitive linguistic system seems to be involved in speakers’ conceptualizations of idioms: the conceptual system. While the conceptual system interfaces with the semantics of an idiom, it can be viewed as a distinct cognitive component for linguistic semantics, evidenced by the fact that a concept can exist to a speaker without a clear sense of expression through language.

Gibbs (1995) provides a thorough account of what it means for an idiom to be “analyzable” from a conceptual point of view. He particularly argues against the notion that idioms are frozen metaphors, meaning that they historically originated as metaphors but have come to be essentially arbitrary phrase-meaning pairs in a speaker’s synchronic lexicon. He cites a variety of studies that favor semantic analyzability for idioms. Even the patterns used to argue for decomposability show that speakers can analyze many idioms and understand their metaphorical motivation. More importantly, experimentation has established that speakers have reliable intuitions regarding the details of the metaphors involved in many idioms. For example, Gibbs and O’Brien (1990) found that speakers can effortlessly answer the following questions reliably for *spill the beans*: “Where are the beans before they are spilled? . . . What caused the beans to spill? . . . After the beans are spilled, are they easy to retrieve?” Combining this metaphorical evidence with the evidence from more traditional decomposability-based studies (many from his own lab), Gibbs (1995) summarizes, “We may not be able to predict exactly what an idiom means through an analysis of the meanings of its individual words, but we can do more than throw our hands up and simply assert that the meanings of idioms are arbitrary and noncompositional.”

Not only do idioms connect to underlying metaphors for which speakers have clear intuitions, many idioms exist in families connected to the same metaphor. For example, *spill the beans* connects to the underlying “conceptual metaphors” THE MIND IS A CONTAINER and IDEAS ARE PHYSICAL ENTITIES (Lakoff and Johnson, 1980), metaphors which it shares with other idioms such as *let the cat out of the bag*. The existence of such underlying metaphors helps to explain the cross-linguistic existence of families of idioms meaning similar things (Gibbs, 1980). For example we can see many idioms related to anger such as *blow your stack*, *flip your lid*, *hit the ceiling*, *get hot under the collar*, etc. behind which underlies the conceptual metaphor ANGER IS HEATED FLUID IN A CONTAINER, which could break down into ANGER IS PRESSURIZED HEAT and THE MIND IS A CONTAINER. The existence of such metaphors has been corroborated experimentally based on participants’ appropriateness ratings for idioms based on whether the context facilitates the metaphor, and even our general intuitions confirm the idea that these metaphors exist. For example, the literal paraphrases we generally assign to idioms do not seem to fully capture the force of the meaning. Although we might define *blow your stack* as ‘get very angry,’ such a definition does not entirely capture the meaning of *blow your stack*, because it excludes the vivid underlying metaphor. For a more in-depth overview of conceptual metaphors and the evidence for their validity, see Gibbs (1995).

Conceptual metaphors are particularly relevant to this thesis because they have already been shown to influence speakers' judgments of the flexibility of idioms. In proposing the importance of conceptual metaphors, Gibbs (1995) particularly points to studies showing that the analyzability of idioms leads to greater lexical flexibility (Gibbs et al., 1989). Unfortunately, these studies performed analysis primarily according to decomposability, but my data provide an avenue to examine this in greater detail. While this line of approach looks promising, the complexities of many-to-many mappings at two different interfaces (syntax-semantics and semantics-conceptual) renders precise accounts nearly impossible. However, thinking of these interfaces in terms of the lexicon can help us to begin to understand how analyzability and flexibility come together.

4.3 The Lexicon and Flexibility

Gibbs (1995) argues that the existence of “coherent systems of metaphorical concepts” (i.e. the “conceptual metaphors” described in Section 4.2.1) contradicts the hypothesis that idioms exist as units in the lexicon. This strikes me as an unnecessary logical leap. Accepting that speakers of English (and perhaps all humans) possess in their conceptual structure a certain set of conceptual metaphors such as ANGER IS HEATED FLUID IN A CONTAINER. This might explain historically why so many of these metaphors exist and why they “catch on.” However, this does not explain how speakers come to use exactly the same phrases based on this metaphor in lieu of other logical possibilities. This issue manifests itself in two ways: the relative lack of occurrence of novel phrases based on this metaphor (47b), and the existence of phrases that, while easily interpreted in relation to this cognitive metaphor, would sound quaint or archaic if used to describe literal heated fluid in a container (48). Indeed, I find it impossible even to construct a literal sentence containing the idiom *blow your stack*.

- (47) a. When Peter found out that his girlfriend was cheating on him, he **flipped his lid**.
 b. ?#When Peter found out that his girlfriend was cheating on him, he **burst his seal**.
- (48) ?#The air in the container was so pressurized that it **flipped its lid**.

In other words, just because the metaphorical machinery of the conceptual structure accommodates an understanding of the idiom without prior knowledge does not mean that the idiom is not stored. We could put this rhetorically, as Jackendoff (1995) does: if the multi-word-items that are ubiquitous to English are not lexical, what are they? The psychological literature, including the work of Gibbs and colleagues, as well as Tabossi and colleagues, seems to miss the linguistic point here. The phrases must be stored in order to become conventional, regardless of whether they can be understood by virtue of knowing the constituent words and the underlying conceptual metaphor. The mental lexicon provides the machinery to account for this.

Of course, this raises redundancy issues. We want to say the the words in any conventionalized

phrase “are” the words that exist alone. We can motivate this through example sentences alone. Irregular verb morphology is preserved in idiomatic constructions (49), and even in nondecomposable idioms, some aspectual properties of the verb remain (50).

- (49) I **took**/***taked** him to the cleaners.
- (50) a. He **lay dying** for weeks.
 b. #He **lay kicking the bucket** for weeks.

Psycholinguistic evidence also strongly supports the notion that the standalone lexical items actually comprise the idioms. This has been most robustly shown for verbs in German VP idioms, where the literal meaning of the verb is activated along with the figurative meaning of the phrase at a 500ms delay, appearing after Cacciari and Tabossi’s (1988) “idiom key” (Smolka et al., 2007; Rabanus et al., 2008). This suggests that at literal meaning of the verb is important to processing, a finding that provides online corroboration to many of our previous points: the compositional use of light verb expressions, the general existence of verb + idiom collocations, and the preservation of aspectual properties observed in (50).

However, this evidence can permit the following line of reasoning *against* a lexical status of many “idioms”: the lexicon contains the form-meaning correspondences that need to be stored in order to be understood. Many idioms can be understood by combining the literal words compositionally in a way that connects to pre-existing conceptual metaphors. Therefore, these idioms are compositional, not lexical. As far as I can tell, this is the approach of the primary researchers who have examined idioms from the perspective of psychology, such as Gibbs (1995) and Tabossi et al. (2009a). This approach makes two unnecessary assumptions: (1) the lexicon does not contain information that can be derived compositionally, and (2) a lexical item cannot be internally analyzed. Even if Jackendoff’s (1997) proposal overplays the importance of decomposability to the lexicon, all of his work regarding the phrasal lexicon demonstrates that these assumptions are invalid (Jackendoff, 1995, 1997, 2002). He points out that redundancy already needs to exist in morphology (Jackendoff, 1995), and also appeals to semantic inheritance networks as a possible explanation (Jackendoff, 2002).

In short, the issue of whether idioms should be considered lexical purely hinges on how we want to define lexical items. In my view, we can even accommodate the idea of lexical items containing other lexical items, which is what Tabossi et al. (2009b) calls a “configuration.” Thus, we know that one component of knowing how to use an idiom properly is simply having a representation of the words in the idiom mapped to the meaning of the idiom as a whole. Regardless of what we call the stored unit necessary for the existence of any formulaic language, we have not yet explained the source of flexibility. The results from Experiment 1 show that most of the cleanly-defined principles that have been proposed to explain idiom flexibility prove to have only limited predictive value, particularly the notion of decomposability. Thus, Jackendoff’s (1997) lexical account of decomposability cannot adequately capture the data. At the same time, the Configuration Hypothesis offers no actual

explanation of the flexibility data. Tabossi et al. (2009b) simply argue that idiom syntax is not idiosyncratic, therefore “general syntactic principles” must operate over “configurations.”

This assertion fails to account for the most basic observations of idiom flexibility, that literally well formed nondecomposable idioms like the infamous *kick the bucket* can accept passivization in a literal context (51a), but the idiomatic reading becomes unavailable (51b).

- (51) a. A: How come the bucket is on its side?
 B: The bucket was kicked by the old man last night.
- b. A: Is the old man still hanging in there?
 B: #No, the bucket was kicked by the old man last night.

This suggests that at least some of the rules governing idiom flexibility must be specific to idioms. Here, the notion of analyzability, particularly with respect to underlying metaphors, becomes particularly important. Existing proposals treat this as a syntactic issue, explained by the syntax-semantics interface. Alternately, we could treat idiom flexibility as an issue of interpretation: can a speaker make sense of a modified idiom in comparison to its canonical form? Under this view, it makes sense that many factors would enter the equation. In the case of substitutions, the clarity of the relationship between the original and substituted word, as well as their association strength, would clearly enter the equation. While this is difficult to quantify, this seemed to play a role in speakers’ judgments of substitutions in Experiment 1. Additionally, inserting an adjective only works if the the new meaning is understandable.

This proposal requires a view of the lexicon that interfaces with the conceptual structure. A modified idiom makes sense if the new version has an understandable relationship to the underlying conceptual form. This means that flexibility occurs at the semantics-conceptual interface, which then interfaces with syntax. Thus, we need to allow that the semantic portion of a lexical representation connects the conceptual system, bridging the gap between the the formal properties of the utterance and its more abstract meaning. This sketch of a proposal is fully consistent with the diversity of flexibility that we actually have observed, but because it predicts messy patterns, it is difficult to test. On a broad scale, we might predict that idioms with the clearest or most common conceptual metaphors would exhibit the greatest flexibility. My data could permit such a measure, if a property like “clarity of metaphor” could be measured for each idiom. If anything, the “meaningfulness” and “literality” measurement most closely approximate this, and together, these factors significantly interacted with all types of condition.

4.4 Conclusion

In this thesis, we are seeking an answer to the question of what speakers need to know in order to properly use an idiom. We have found that speakers require something like a lexical item. This

lexical item might be embedded in a collocation, and regardless, syntactic flexibility is not specified as part of an idiom's lexical item. Moreover, the meaning of an idiom is likely represented in a clear way, judging by the fact that very few items receive different decomposability ratings simply based on how the experiment approximates their meaning.

Since syntactic flexibility is not lexical, speakers must appear to principled linking rules in order to determine the flexibility of idioms. The data from both experiments move us toward a model where idioms (and collocations) are stored in the lexicon with a complex set of linking rules between levels of their representation – syntactic, semantic, and conceptual. Previously proposed principles of linking, such as decomposability, can sometimes approximate this, but decomposability is only a part of a larger picture of internal analysis and the ability to recover meaning from modified forms of idioms. Truly, these complex lexical entries and rules of correspondence across representational levels constitute a speaker's knowledge of idioms. The final chapter reviews my findings, their implications, and directions for further research.

Chapter 5

Wrapping Up

I return to the question I posed at the beginning of this thesis: what does a speaker need to know in order to properly use an idiom? We have sought answers to this question by asking how (or if) idioms are represented in the lexicon and what principles explain their syntactic and lexical flexibility, focusing particularly on the idea of idiom decomposability.

We began by examining the claims of the Idiom Decomposability Hypothesis (Gibbs and Nayak, 1989) and the various lexical representations associated therewith. In Experiment 1, we used a diverse set of stimuli and new manipulations to idioms to determine that speakers' judgments of flexibility are just as consistent for idioms they know as they are for novel idioms. This tells us that the syntactic flexibility of idioms cannot be prespecified as part of their lexical representation. Therefore, knowing how an idiom is properly used requires more than just having stored the right information about the idiom. Indeed, some factors such as decomposability provide a partial explanation for some modifications but not others, indicating that flexibility is a multidimensional scale for idioms.

In Experiment 2, we confirmed that speakers usually agree with experimenters when it comes to the sorting of idioms into the bins of “decomposable” and “nondecomposable.” They generally perform this sorting at an equal rate regardless of how the idiom’s meaning is paraphrased to them. This tells us that speakers have clear intuitions about the meaning of idioms, ones that cannot easily be perturbed. This suggests that part of competent use and understanding of idioms is having a clear mental representation of their meaning.

We concluded our analysis by looking at items, understanding that idioms must be represented in the lexicon, but this provides only occasional explanation for flexibility. Particularly, the lexical representation of an idiom predicts its flexibility when the lexical item is a collocation containing an idiom. Beyond this, we looked more abstractly to relations between or conceptualization of an idiom and its linguistic structure. This is a level of abstraction above ideas like decomposability. It does not lend itself to immediate, clean formalizations, but rather suggests a large area of future

research where speakers' ability to analyze and visualize idioms influences their semantic structure and syntactic flexibility.

I now conclude this thesis by saying what we know is *not* the case, what we know is the case, and what we have yet to know. We know that a given idiom's flexibility is not stored idiosyncratically in the lexicon or some equivalent, since in comparison to known idiom controls, participants have equally reliable judgments for idioms they have heard before as idioms they have not. Moving to principles, we know from Experiment 1 that speakers do not readily accept scalar substitution. We also know that decomposability does not cleanly explain the entirety of idiom flexibility and that idiom flexibility is not a singular concept; an idiom can be differently flexible with different modifications or different contexts. We learned this through linear modeling of acceptability ratings for a representative set of normed idioms.

What do we know is the case? Idioms must be represented in the mind by definition, regardless of whether we choose to label this representation a lexical item or a "configuration." These representations connect the morphology and phonology of the constituent words, the global meaning, and often, an underlying conceptual metaphor. The results of Experiment 2 suggest that, barring particularly biasing examples, speakers retain a steady understanding of an idiom's meaning, rather than being easily persuaded to reconceptualize it. This further establishes that idiom meanings are stored, rather than being (close to) compositional.

Rather than idiosyncratically representing flexibility through lexical storage, idioms pattern according to some principles. Decomposability does offer some explanatory power, but its very weak correlation with only some types of modifications shows that it is only a fragment of the picture. The judgment data from Experiment 1 showed that we can extend the idea of analyzing the idiom further than the original definition of decomposability. In some of the greatest cases of flexibility, the subparts of the "idiom," such as *cost an arm and a leg* can be reanalyzed as a smaller idiom (*arm and a leg*) embedded in a more compositionally-understood collocation. In the case of light verb idioms, nearly all of the idiomatic meaning is contained in one word (generally, the noun), but this requires the context of a light verb expression in order to be understood. These examples do not correspond exactly to the original formal statement of decomposability, but they very much follow the spirit of understanding an idiom's parts. Beyond this, the idea of conceptual metaphors and previous work presented by Gibbs (1995) show that speakers have clear intuitions about the metaphors underlying certain sets of idioms; this knowledge of the "source" of an idiom can help a speaker determine whether an idiom is flexible. Indeed, my own results showed that speakers were more willing to accept modifications to plausible novel idioms, presumably because speakers could make sense of the inner workings of the idioms.

These results leave many questions unanswered. Conceptual metaphors begin to take us into murky waters outside of the domain of linguistics, but future cognitive scientists will hopefully begin to characterize the interplay between idiom flexibility and conceptual metaphors. Researchers could start by creating a metric for conceptual metaphor strength, which could be used against my

data to determine its predictive value. Generally, however, we have seen that many principles are likely involved in idiom flexibility, and these principles probably exist at the interfaces among syntax, semantics, pragmatics, and conceptual structure. As we uncover these principles, we can discover more about other open questions, such as the relationship between the lexical items for individual words in an idiom and the idiom's own lexical representation. Finally, speakers do vary in their ability to accept idiom modifications; understanding the source of this variation could be a worthwhile future pursuit.

In conclusion, speakers need both a lexical representation of an idiom and principles to understand it internally in order to properly use an idiom. Here, I have broadened our data set, looked at underexamined items, and systematically checked many representative idioms for a series of manipulations. I have opened the can of worms, documented a few new specimens, and taken a broader look at all of the contents of the can. As we continue to dig through the can of worms I have opened here, I trust that researchers will find an enriched understanding of the human language faculty.

Appendix A

Experiment 1 Stimuli

A.1 Real

A.1.1 VP

1. to give someone the cold shoulder (to ignore someone)
 - (a) **Control:** After Janice insulted him deeply, Ricky gave her the cold shoulder for weeks.
 - (b) **Adjective Insertion:** After Janice insulted him deeply, Ricky gave her the cold of-fended shoulder for weeks.
 - (c) **Scalar Substitution:** Janice insulted Ricky so deeply that he gave her the frigid shoul-der for weeks.
 - (d) **Non-scalar Substitution:** After Janice insulted him deeply, Ricky gave her the chilly shoulder for weeks.
 - (e) **Passive:** After Janice insulted him deeply, she was given the cold shoulder by Ricky for weeks.
2. to slip someone's mind (to forget something)
 - (a) **Control:** Erica meant to stop by the post office, but it slipped her mind.
 - (b) **Adjective Insertion:** Erica meant to stop by the post office, but it slipped her hectic mind.
 - (c) **Scalar Substitution:** Erica meant to stop by the post office, but the memory of her afternoon tasks actively evaded her mind.
 - (d) **Non-scalar Substitution:** Erica meant to stop by the post office, but it slipped her brain.
 - (e) **Passive:** Erica meant to stop by the post office, but her mind was slipped by it.
3. to get the picture (to understand what's going on)
 - (a) **Control:** By the time we finished explaining the conflict, Evan got the picture.

- (b) **Adjective Insertion:** By the time we finished explaining the conflict, Evan got the ugly picture.
 - (c) **Scalar Substitution:** We explained the conflict so thoroughly that by the time we were done, Evan got the whole gallery.
 - (d) **Non-scalar Substitution:** By the time we finished explaining the conflict, Evan got the image.
 - (e) **Passive:** By the time we finished explaining the conflict, the picture was gotten by Evan.
4. to call the shots (to make the decisions)
- (a) **Control:** Margaret is the real boss around here; she calls the shots.
 - (b) **Adjective Insertion:** Margaret is the real boss around here; she calls the important shots.
 - (c) **Scalar Substitution:** Margaret is an extremely forceful boss; she yells the shots.
 - (d) **Non-scalar Substitution:** Margaret is the real boss around here; she hollers the shots.
 - (e) **Passive:** Margaret is the real boss around here; the shots are called by her.
5. to face the music (to accept the unpleasant results of one's actions)
- (a) **Control:** Even though he was afraid of the grave repercussions of his actions, Henry decided to face the music and confess to the crime.
 - (b) **Adjective Insertion:** Even though he was afraid of the grave repercussions of his actions, Henry decided to face the legal music and confess to the crime.
 - (c) **Scalar Substitution:** Even though he was afraid of the grave repercussions of his actions, Henry decided to face the symphony and confess to the crime.
 - (d) **Non-scalar Substitution:** Even though he was afraid of the grave repercussions of his actions, Henry decided to confront the music and confess to the crime.
 - (e) **Passive:** Even though Henry was afraid of the repercussions of his actions, the music was faced by him when he confessed to the crime.
6. to hit the sack (to go to bed)
- (a) **Control:** Hannah's day at work was so tiring that she hit the sack as soon as she arrived at home.
 - (b) **Adjective Insertion:** Hannah's day at work was so tiring that she hit the cozy sack as soon as she arrived at home.
 - (c) **Scalar Substitution:** Hannah's day at work was very tiring, but since she had evening plans, she could only touch the sack when she arrived at home.
 - (d) **Non-scalar Substitution:** Hannah's day at work was so tiring that she hit the bag as soon as she arrived at home.
 - (e) **Passive:** Hannah's day at work was so tiring that the sack was hit by her as soon as she arrived at home.

7. to swallow one's pride (to accept something humiliating despite pride)
 - (a) **Control:** When he realized he was in the wrong for their long debate, Cameron swallowed his pride and apologized to Susan.
 - (b) **Adjective Insertion:** When he realized he was in the wrong for their long debate, Cameron swallowed his argumentative pride and apologized to Susan.
 - (c) **Scalar Substitution:** When he realized he was profoundly in the wrong for their long debate, Cameron completely digested his pride and apologized to Susan.
 - (d) **Non-scalar Substitution:** When he realized he was in the wrong for their long debate, Cameron swallowed his arrogance and apologized to Susan.
 - (e) **Passive:** When Cameron realized he was in the wrong for their long debate, Cameron's pride was swallowed by him, and he apologized to Susan.
8. to jump the gun (to start something too soon)
 - (a) **Control:** When Alyssa implemented solutions to the problem without hearing everyone out, Joel complained that she had jumped the gun.
 - (b) **Adjective Insertion:** When Alyssa implemented solutions to the problem without hearing everyone out, Joel complained that she had jumped the administrative gun.
 - (c) **Scalar Substitution:** When Alyssa implemented solutions to the problem without hearing everyone out, Joel complained that she had completely leaped the gun.
 - (d) **Non-scalar Substitution:** When Alyssa implemented solutions to the problem without hearing everyone out, Joel complained that she had jumped the pistol.
 - (e) **Passive:** When Alyssa implemented solutions to the problem without hearing everyone out, Joel complained that the gun had been jumped by her.
9. to learn by heart (to memorize)
 - (a) **Control:** Katie's memorization skills were impressive; there were fifteen violin concertos that she had learned by heart.
 - (b) **Adjective Insertion:** Katie's memorization skills were impressive; there were fifteen violin concertos that she had learned by musical heart.
 - (c) **Scalar Substitution:** Katie's had a remarkable ability to internalize music; there were fifteen violin concertos that she had learned by soul.
 - (d) **Non-scalar Substitution:** Katie's memorization skills were impressive; there were fifteen violin concertos that she had remembered by heart.
 - (e) **Passive:** Katie's memorization skills were impressive; there were fifteen violin concertos that had been learned by heart by her.
10. to twist someone's arm (to pressure someone)
 - (a) **Control:** Bill didn't want to go to the party, but he agreed to go after Jessie twisted his arm.
 - (b) **Adjective Insertion:** Bill didn't want to go to the party, but he agreed to go after Jessie twisted his resisting arm.

- (c) **Scalar Substitution:** It didn't take much to convince Bill to go to the party; Jessie only needed to squeeze his arm before he agreed to go.
 - (d) **Non-scalar Substitution:** Bill didn't want to go to the party, but he agreed to go after Jessie contorted his arm.
 - (e) **Passive:** Bill didn't want to go to the party, but he agreed to go after his arm was twisted by Jessie.
11. break the ice (to start a conversation)
- (a) **Control:** At the awkward cocktail party, Harry told a great joke in order to break the ice.
 - (b) **Adjective Insertion:** At the awkward cocktail party, Harry told a great joke in order to break the stuffy ice.
 - (c) **Scalar Substitution:** At the awkward cocktail party, Harry told a great joke, totally changing the atmosphere and shattering the ice.
 - (d) **Non-scalar Substitution:** At the awkward cocktail party, Harry told a great joke in order to split the ice.
 - (e) **Passive:** At the awkward cocktail party, Harry told a great joke, and the ice was broken by him.
12. cost an arm and a leg (to be very expensive)
- (a) **Control:** Gas is twice as expensive as it was ten years ago; now, it really costs an arm and a leg!
 - (b) **Adjective Insertion:** Gas is twice as expensive as it was ten years ago; now, it really costs an arm and a pricey leg!
 - (c) **Scalar Substitution:** Gas is twice as expensive as it was ten years ago; now, it really costs both arms and a leg!
 - (d) **Non-scalar Substitution:** Gas is twice as expensive as it was ten years ago; now, it really sells for an arm and a leg!
 - (e) **Passive:** Gas is twice as expensive as it was ten years ago; now, an arm and a leg is cost by it!
13. lose one's grip (to lose one's ability to deal with a situation) [Eliminated from analysis]
- (a) **Control:** The elderly professor's teaching skills had worsened drastically over the past decade; he decided to retire, since he was losing his grip.
 - (b) **Adjective Insertion:** The elderly professor's teaching skills had worsened drastically over the past decade; he decided to retire, since he was losing his mental grip.
 - (c) **Scalar Substitution:** The elderly professor's teaching skills had worsened drastically over the past decade; he decided to retire, since he was losing his grip.
 - (d) **Non-scalar Substitution:** The elderly professor's teaching skills had worsened drastically over the past decade; he decided to retire, since he was losing his grasp.

- (e) **Passive:** The elderly professor’s teaching skills had worsened drastically over the past decade; he decided to retire, since his grip was being lost by him.
14. play with fire (to do something risky)
- (a) **Control:** When Joe joined a gang in order to feel protected, he was really playing with fire.
- (b) **Adjective Insertion:** When Joe joined a gang in order to feel protected, he was really playing with criminal fire.
- (c) **Scalar Substitution:** Sometimes hanging out with criminals in order to be protected is kind of dabbling with fire.
- (d) **Non-scalar Substitution:** When Joe joined a gang in order to feel protected, he was really playing with flames.
- (e) **Passive:** When Joe joined a gang in order to feel protected, fire was being played with by him.
15. tie the knot (to get married)
- (a) **Control:** Kevin and Alice had been everyone’s favorite couple for four years, so no one was surprised when they finally tied the knot.
- (b) **Adjective Insertion:** Kevin and Alice had been everyone’s favorite couple for four years, so no one was surprised when they finally tied the traditional knot.
- (c) **Scalar Substitution:** Kevin and Alice had been everyone’s favorite couple for four years, so no one was surprised when they finally completely secured the knot.
- (d) **Non-scalar Substitution:** Kevin and Alice had been everyone’s favorite couple for four years, so no one was surprised when they finally bound the knot.
- (e) **Passive:** Kevin and Alice had been everyone’s favorite couple for four years, so no one was surprised when the knot was finally tied by them.

A.1.2 PP

1. out of the blue (sudden(ly) or unexpected(ly))
- (a) **Control:** Liz hadn’t heard from Emily in years, but right before their class reunion, Emily sent her an email out of the blue.
- (b) **Adjective Insertion:** Liz hadn’t heard from Emily in years, but right before their class reunion, Emily sent her an email out of the social blue.
- (c) **Scalar Substitution:** Liz hadn’t heard from Emily in decades, but right before their class reunion, Liz was shocked to receive an email from her out of the ultraviolet.
- (d) **Non-scalar Substitution:** Liz hadn’t heard from Emily in years, but right before their class reunion, Emily sent her an email out of the azure.
2. out of thin air (out of nowhere)

- (a) **Control:** Patrick knew many high-profile band managers, so it didn't surprise us when he got front-row tickets to the sold out concert out of thin air.
 - (b) **Adjective Insertion:** Patrick knew many high-profile band managers, so it didn't surprise us when he got front-row tickets to the sold out concert out of thin clear air.
 - (c) **Scalar Substitution:** Patrick didn't seem well connected, so we were shocked when he got front-row tickets to the sold out concert out of meager air.
 - (d) **Non-scalar Substitution:** Patrick knew many high-profile band managers, so it didn't surprise us when he got front-row tickets to the sold out concert out of light air.
3. over the hill (past prime)
- (a) **Control:** The 40-year-old quarterback decided to retire since he had been over the hill for a long time.
 - (b) **Adjective Insertion:** The 40-year-old quarterback decided to retire since he had been over the athletic hill for a long time.
 - (c) **Scalar Substitution:** The 40-year-old quarterback decided to retire since, in comparison to the outstanding pinnacle of his career, he had been over the mountain for a long time.
 - (d) **Non-scalar Substitution:** The 40-year-old quarterback decided to retire since he had been past the hill for a long time.
4. in hot water (in a difficult situation likely leading to punishment)
- (a) **Control:** When Kelly's boss found out that she spent most of her time at work shopping online, Kelly was really in hot water.
 - (b) **Adjective Insertion:** When Kelly's boss found out that she spent most of her time at work shopping online, Kelly was really in hot occupational water.
 - (c) **Scalar Substitution:** When Kelly's hot-tempered boss found out that she spent most of her time at work shopping online, Kelly was really in boiling water.
 - (d) **Non-scalar Substitution:** When Kelly's boss found out that she spent most of her time at work shopping online, Kelly was really in heated water.
5. in the nick of time (just in time)
- (a) **Control:** Justin got caught in traffic and thought he would miss his flight, but he arrived in the nick of time.
 - (b) **Adjective Insertion:** Justin got caught in traffic and thought he would miss his flight, but he arrived in the nick of traveling time.
 - (c) **Scalar Substitution:** Justin got caught in traffic for many hours and thought he would miss his flight, but he arrived in the nick of eons.
 - (d) **Non-scalar Substitution:** Justin got caught in traffic and thought he would miss his flight, but he arrived in the chip of time.
6. with flying colors (easily, with great success)

- (a) **Control:** After studying hard for several months, the brilliant law student passed the bar exam with flying colors.
 - (b) **Adjective Insertion:** After studying hard for several months, the brilliant law student passed the bar exam with flying bright colors.
 - (c) **Scalar Substitution:** After studying hard for several months, the brilliant law student passed the bar exam with soaring colors.
 - (d) **Non-scalar Substitution:** After studying hard for several months, the brilliant law student passed the bar exam with flying pigments.
7. under the weather (sick/ill)
- (a) **Control:** Stephen wasn't able to attend the meeting since he was feeling under the weather.
 - (b) **Adjective Insertion:** Stephen wasn't able to attend the meeting since he was feeling under the viral weather.
 - (c) **Scalar Substitution:** Now that he works out daily, Stephen constantly claims that he is feeling over the weather.
 - (d) **Non-scalar Substitution:** Stephen wasn't able to attend the meeting since he was feeling under the elements.
8. in seventh heaven (very happy)
- (a) **Control:** When Alex's parents bought him a new luxury car for Christmas, he was really in seventh heaven.
 - (b) **Adjective Insertion:** When Alex's parents bought him a new luxury car for Christmas, he was really in seventh driving heaven.
 - (c) **Scalar Substitution:** When Alex's parents bought him a new luxury car for Christmas, he was really in eighth heaven.
 - (d) **Non-scalar Substitution:** When Alex's parents bought him a new luxury car for Christmas, he was really in seventh paradise.
9. at the back of one's mind (vaguely remembered)
- (a) **Control:** Jesse was very focused on searching for a job, so she kept her relationship's impending failure at the back of her mind.
 - (b) **Adjective Insertion:** Jesse was very focused on searching for a job, so she kept her relationship's impending failure at the back of her anxious mind.
 - (c) **Scalar Substitution:** Jesse was very focused on searching for a job, but she couldn't help but keep her relationship's impending failure at the middle of her mind.
 - (d) **Non-scalar Substitution:** Jesse was very focused on searching for a job, so she kept her relationship's impending failure at the rear of her mind.
10. on thin ice (in a risky/precarious situation)
- (a) **Control:** When the athlete sprained his knee in a drunken accident, he was on thin ice with the coach.

- (b) **Adjective Insertion:** When the athlete sprained his knee in a drunken accident, he was on thin playing ice with the coach.
- (c) **Scalar Substitution:** When the athlete sprained his knee after his fourth drunken accident, he was on cracking ice with the coach.
- (d) **Non-scalar Substitution:** When the athlete sprained his knee in a drunken accident, he was on weak ice with the coach.

A.1.3 NP

1. food for thought (something to ponder)
 - (a) **Control:** The professor asked her students a complex ethical question, but she didn't provide an answer; the question was just food for thought.
 - (b) **Adjective Insertion:** The professor asked her students a complex ethical question, but she didn't provide an answer; the question was just food for careful thought.
 - (c) **Scalar Substitution:** The professor asked her students a complex ethical question, and she allowed them a few weeks to ponder it; the question was an entire meal for thought.
 - (d) **Non-scalar Substitution:** The professor asked her students a complex ethical question, but she didn't provide an answer; the question was just food for consideration.
2. rule of thumb (a general principle)
 - (a) **Control:** The driving instructor suggested that one car length for every 10 mph was a good rule of thumb.
 - (b) **Adjective Insertion:** The driving instructor suggested that one car length for every 10 mph was a good rule of driving thumb.
 - (c) **Scalar Substitution:** The driving instructor suggested that one car length for every 10 mph was such a good idea that Joey should consider it a law of thumb.
 - (d) **Non-scalar Substitution:** The driving instructor suggested that one car length for every 10 mph was a good policy of thumb.
3. tip of the iceberg (part of something much bigger)
 - (a) **Control:** In the major legal case against the financial company, tax evasion was just the tip of the iceberg.
 - (b) **Adjective Insertion:** In the major legal case against the financial company, tax evasion was just the tip of the corrupt iceberg.
 - (c) **Scalar Substitution:** In the major legal case against the financial company, tax evasion was pretty important; it was the top half of the iceberg.
 - (d) **Non-scalar Substitution:** In the major legal case against the financial company, tax evasion was just the point of the iceberg.
4. a frog in one's throat (a state of anxiousness)

- (a) **Control:** Any time that he was about to speak in front of a large crowd, Philip got a frog in his throat.
 - (b) **Adjective Insertion:** Any time that he was about to speak in front of a large crowd, Philip got a frog in his nervous throat.
 - (c) **Scalar Substitution:** Any time that he was about to speak in front of a large crowd, Philip gets so nervous that he has an entire swamp in his throat.
 - (d) **Non-scalar Substitution:** Any time that he was about to speak in front of a large crowd, Philip got a toad in his throat.
5. fish out of water (someone who is in a situation he or she is unsuited to)
- (a) **Control:** The clumsy girl felt like a fish out of water next to the skilled dancers at her Homecoming Dance.
 - (b) **Adjective Insertion:** The clumsy girl felt like a fish out of motionless water next to the skilled dancers at her Homecoming Dance.
 - (c) **Scalar Substitution:** The clumsy girl felt like a whale out of water next to the skilled dancers at her Homecoming Dance.
 - (d) **Non-scalar Substitution:** The clumsy girl felt like a fish out of the ocean next to the skilled dancers at her Homecoming Dance.
6. two peas in a pod (two of a kind)
- (a) **Control:** Those boys running around the buffet were really two peas in a pod.
 - (b) **Adjective Insertion:** Those boys running around the buffet were really two gluttonous peas in a pod.
 - (c) **Scalar Substitution:** Those boys running around the buffet were really a large group of peas in a pod.
 - (d) **Non-scalar Substitution:** Those boys running around the buffet were really two gluttonous peas in a pouch.
7. par for the course (something expected or normal)
- (a) **Control:** Amy didn't get into her top choice for college, but that was just par for the course at her inner-city school.
 - (b) **Adjective Insertion:** Amy didn't get into her top choice for college, but that was just par for the difficult course at her inner-city school.
 - (c) **Scalar Substitution:** Amy didn't get into her top choice for college, and at her elite prep school such failure was bogey for the course.
 - (d) **Non-scalar Substitution:** Amy didn't get into her top choice for college, but that was just average for the course at her inner-city school.
8. letter of the law (the literal interpretation of regulations) [Due to the typo in the word *stuck*, this was eliminated from analysis.]
- (a) **Control:** George was a major stickler to the speed limit; no matter what, he sucked to the letter of the law.

- (b) **Adjective Insertion:** George was a major stickler to the speed limit; no matter what, he suck to the letter of the stringent law.
 - (c) **Scalar Substitution:** George was a major stickler to the speed limit; no matter what, he stuck to every penstroke of the law.
 - (d) **Non-scalar Substitution:** George was a major stickler to the speed limit; no matter what, he suck to the letter of the rules.
9. a piece of cake (something very easy to do)
- (a) **Control:** In comparison to running a marathon, running a few laps around the track is a piece of cake.
 - (b) **Adjective Insertion:** In comparison to running a marathon, running a few laps around the track is a piece of athletic cake.
 - (c) **Scalar Substitution:** Even after running a marathon, a 5k sprint is a sheet of cake.
 - (d) **Non-scalar Substitution:** In comparison to running a marathon, running a few laps around the track is a slice of cake.
10. a horse of another color (another matter altogether)
- (a) **Control:** Gambling is not the same as investing in the stock market; it's a horse of another color.
 - (b) **Adjective Insertion:** Gambling is not the same as investing in the stock market; it's a horse of another foolish color.
 - (c) **Scalar Substitution:** Gambling is extremely different from investing in the stock market; it's an entire herd of another color.
 - (d) **Non-scalar Substitution:** Gambling is not the same as investing in the stock market; it's a pony of another color.

A.2 Novel Idioms

A.2.1 VP

1. to roll hat (to party)
- (a) **Control:** The neighbors of the fraternity house called the police due to the noise the students made as they rolled hat.
 - (b) **Adjective Insertion:** The neighbors of the fraternity house called the police due to the noise the students made as they rolled festive hat.
 - (c) **Scalar Substitution:** The neighbors of the fraternity house called the police due to the intense amount of noise the students made as they wildly spun cap.
 - (d) **Non-scalar Substitution:** The neighbors of the fraternity house called the police due to the noise the students made as they rolled cap.

- (e) **Passive:** The neighbors of the fraternity house called the police due to the noise the students made as hat was rolled by them.
2. to walk on the pump (to try something and fail)
- (a) **Control:** Harry took dance lessons, but in the end he couldn't even sway to the beat; he really walked on the pump.
- (b) **Adjective Insertion:** Harry took dance lessons, but in the end he couldn't even sway to the beat; he really walked on the rhythmic pump.
- (c) **Scalar Substitution:** Harry took dance lessons, but in the end he couldn't even sway to the beat; he was so bad that he stomped on the pump.
- (d) **Non-scalar Substitution:** Harry took dance lessons, but in the end he couldn't even sway to the beat; he really strolled on the pump.
- (e) **Passive:** Harry took dance lessons, but in the end he couldn't even sway to the beat; the pump was walked on by him.
3. to throw the spoon (to die)
- (a) **Control:** Considering his crotchety demeanor and heavy smoking, it was about time that the elderly man threw the spoon.
- (b) **Adjective Insertion:** Considering his crotchety demeanor and heavy smoking, it was about time that the elderly man threw the diseased spoon.
- (c) **Scalar Substitution:** Considering his crotchety demeanor and heavy smoking, it was surprising that the old man quietly tossed the spoon.
- (d) **Non-scalar Substitution:** Considering his crotchety demeanor and heavy smoking, it was about time that the elderly man chucked the spoon.
- (e) **Passive:** Considering his crotchety demeanor and heavy smoking, it was about time that the spoon was thrown by the elderly man.
4. to drop the suck (to give up)
- (a) **Control:** After five years of auditioning for the prestigious orchestra and failing, Louis decided it was time to drop the suck.
- (b) **Adjective Insertion:** After five years of auditioning for the prestigious orchestra and failing, Louis decided it was time to drop the musical suck.
- (c) **Scalar Substitution:** After five years of auditioning for the prestigious orchestra and failing, Louis completely abandoned his musical aspirations and threw down the suck.
- (d) **Non-scalar Substitution:** After five years of auditioning for the prestigious orchestra and failing, Louis decided it was time to release the suck.
- (e) **Passive:** After five years of auditioning for the prestigious orchestra and failing, Louis decided it was time for the suck to be dropped by him.
5. to hit a drill (to urinate)
- (a) **Control:** Ben could hardly hold it in after chugging two liters of soda, so he ran into the woods to hit a drill.

- (b) **Adjective Insertion:** Ben could hardly hold it in after chugging two liters of soda, so he ran into the woods to hit a powerful drill.
 - (c) **Scalar Substitution:** Ben could hardly hold it in after chugging two liters of soda, so he ran into the woods and slammed a drill.
 - (d) **Non-scalar Substitution:** Ben could hardly hold it in after chugging two liters of soda, so he ran into the woods to strike a drill.
 - (e) **Passive:** Ben could hardly hold it in after chugging two liters of soda, so he ran into the woods, and a drill was hit by him.
6. to walk in the roof (to be really pissed off)
- (a) **Control:** When someone dented the fender of her new car then sped away, Jenna really walked in the roof.
 - (b) **Adjective Insertion:** When someone dented the fender of her new car then sped away, Jenna really walked in the low-hanging roof.
 - (c) **Scalar Substitution:** When someone dented the fender of her new car then sped away, Jenna completely ran in the roof.
 - (d) **Non-scalar Substitution:** When someone dented the fender of her new car then sped away, Jenna really strolled in the roof.
 - (e) **Passive:** When someone dented the fender of her new car then sped away, the roof was really walked in by Jenna.
7. to grasp the gallop (to understand what's going on)
- (a) **Control:** Larry was pretty slow on the uptake, so it took lots of explaining before he grasped the gallop.
 - (b) **Adjective Insertion:** Larry was pretty slow on the uptake, so it took lots of explaining before he grasped the complicated gallop.
 - (c) **Scalar Substitution:** Larry was pretty slow on the uptake, so it took lots of explaining before he even loosely held the gallop.
 - (d) **Non-scalar Substitution:** Larry was pretty slow on the uptake, so it took lots of explaining before he gripped the gallop.
 - (e) **Passive:** Larry was pretty slow on the uptake, so it took lots of explaining before the the gallop was grasped by him.
8. to give someone the rake (to shake hands) [Excluded from analysis]
- (a) **Control:** At Nancy's first interview, she was so nervous that she forgot common social courtesy and didn't give the interviewer the rake.
 - (b) **Adjective Insertion:** At Nancy's first interview, she was so nervous that she forgot common social courtesy and didn't give the interviewer the polite rake.
 - (c) **Scalar Substitution:** At Nancy's first interview, she was so nervous that she went overboard and gave the interviewer the full toolshed.
 - (d) **Non-scalar Substitution:** [Incorrect sentence was presented]

- (e) **Passive:** At Nancy's first interview, she was so nervous that she forgot common social courtesy; the rake was not given to the interviewer by her.
9. to twist from wire (to explain in very simple terms)
- (a) **Control:** When Carl didn't understand how to code up a simple flashcard program, the teaching assistant realized that it was time to twist from wire.
- (b) **Adjective Insertion:** When Carl didn't understand how to code up a simple flashcard program, the teaching assistant realized that it was time to twist from simple wire.
- (c) **Scalar Substitution:** When Carl didn't understand how to make his computer do addition, the teaching assistant realized that it was time to twist from raw metal.
- (d) **Non-scalar Substitution:** When Carl didn't understand how to code up a simple flashcard program, the teaching assistant realized that it was time to twist from cable.
- (e) **Passive:** When Carl didn't understand how to code up a simple flashcard program, the teaching assistant realized that it was time for wire to be twisted from by her.
10. to put on someone's fallen cloak (to take over for one's predecessor)
- (a) **Control:** After the president was assassinated, the vice president put on his fallen cloak.
- (b) **Adjective Insertion:** After the president was assassinated, the vice president put on his fallen idealistic cloak.
- (c) **Scalar Substitution:** After the president was assassinated, the vice president went through all of his plans and put on his entire fallen wardrobe.
- (d) **Non-scalar Substitution:** After the president was assassinated, the vice president put on his fallen coat.
- (e) **Passive:** After the president was assassinated, his fallen cloak was put on by the vice president.
11. to begin on a new marble (to start all over again)
- (a) **Control:** After being laid off from her retail job, Jennifer was excited for her new job as a paralegal, where she could begin on a new marble.
- (b) **Adjective Insertion:** After being laid off from her retail job, Jennifer was excited for her new job as a paralegal, where she could begin on a new professional marble.
- (c) **Scalar Substitution:** After being laid off from her job as a paralegal, Jennifer returned to her previous job in retail, where she would begin on an old marble.
- (d) **Non-scalar Substitution:** After being laid off from her retail job, Jennifer was excited for her new job as a paralegal, where she could begin on a fresh marble.
- (e) **Passive:** After being laid off from her retail job, Jennifer was excited for her new job as a paralegal, where a new marble was begun on by her.
12. to play one's mouth (to boast)
- (a) **Control:** The champion snowboarder was known for being very confident in his abilities; after he recieved the gold medal, he really played his mouth in interviews.

- (b) **Adjective Insertion:** The champion snowboarder was known for being very confident in his abilities; after he recieved the gold medal, he really played his profane mouth in interviews.
- (c) **Scalar Substitution:** The champion snowboarder was known for being very confident in his abilities; he was so cocky after he recieved the gold medal that he played his whole face in interviews.
- (d) **Non-scalar Substitution:** The champion snowboarder was known for being very confident in his abilities; after he recieved the gold medal, he really played his lips in interviews.
- (e) **Passive:** The champion snowboarder was known for being very confident in his abilities; after he recieved the gold medal, his mouth was really played by him in interviews.
13. to throw a goat's eye (to take a fast look)
- (a) **Control:** When Jean pointed out the senator in the restaurant, her son threw a goat's eye at him.
- (b) **Adjective Insertion:** When Jean pointed out the senator in the restaurant, her son threw a quick goat's eye at him.
- (c) **Scalar Substitution:** When Jean pointed out the senator in the restaurant, her son was so fascinated that he threw a goat's head at him.
- (d) **Non-scalar Substitution:** When Jean pointed out the senator in the restaurant, her son flung a goat's eye at him.
- (e) **Passive:** When Jean pointed out the senator in the restaurant, a goat's eye was thrown by her son at him.
14. to speak out of the beard (to speak up, speak more clearly)
- (a) **Control:** At the conference, Max needed to be heard across a long table, so he spoke out of the beard.
- (b) **Adjective Insertion:** At the conference, Max needed to be heard across a long table, so he spoke out of the knowledgable beard.
- (c) **Scalar Substitution:** At the conference, Max needed to be heard across a long table in a noisy room, so he yelled out of the beard.
- (d) **Non-scalar Substitution:** At the conference, Max needed to be heard across a long table, so he talked out of the beard.
- (e) **Passive:** At the conference, Max needed to be heard across a long table, so the beard was spoken out of by him.
15. to put onion on the salmon (to make things worse)
- (a) **Control:** When Ashley tried to melt the snow in her driveway by pouring hot water, the water just froze over; she was really putting onion on the salmon.
- (b) **Adjective Insertion:** When Ashley tried to melt the snow in her driveway by pouring hot water on it, the water just froze over; she was really putting onion on the wintry salmon.

- (c) **Scalar Substitution:** After the blizzard, Ashley tried to melt the snow in her whole street by pouring hot water on it, the water just froze over; she was really putting onion on the whole school.
- (d) **Non-scalar Substitution:** When Ashley tried to melt the snow in her driveway by pouring hot water on it, the water just froze over; she was really placing onion on the salmon.
- (e) **Passive:** When Ashley tried to melt the snow in her driveway by pouring hot water on, the water just froze over; onion was being put on the salmon by her.

A.2.2 PP

1. behind the floating (stupid)
 - (a) **Control:** When Dan didn't know who the current president was, the whole class realized he was a bit behind the floating.
 - (b) **Adjective Insertion:** When Dan didn't know who the current president was, the whole class realized he was a bit behind the political floating.
 - (c) **Scalar Substitution:** After Dan correctly predicted the entirety of the new president's cabinet, the whole class realized he was a bit ahead of the floating.
 - (d) **Non-scalar Substitution:** When Dan didn't know who the current president was, the whole class realized he was a bit behind the hovering.
2. on the wallpaper (on the agenda)
 - (a) **Control:** Before Ron agreed to attend the meeting with executives from across the world, he wanted to know what was on the wallpaper.
 - (b) **Adjective Insertion:** Before Ron agreed to attend the meeting with executives from across the world, he wanted to know what was on the international wallpaper.
 - (c) **Scalar Substitution:** Before Ron agreed to attend the meeting with executives from across the world, he wanted to make sure that the controversial issue was off the wallpaper.
 - (d) **Non-scalar Substitution:** Before Ron agreed to attend the meeting with executives from across the world, he wanted to know what was on the sheetrock.
3. in the count's time (just in time)
 - (a) **Control:** In order to increase the dramatic tension, it seems that the rescuer in a movie always arrives in the count's time.
 - (b) **Adjective Insertion:** In order to increase the dramatic tension, it seems that the rescuer in a movie always arrives in the count's perfect time.
 - (c) **Scalar Substitution:** Movies would have much sadder endings if the heroes arrived in the DMV's time.
 - (d) **Non-scalar Substitution:** In order to increase the dramatic tension, it seems that the rescuer in a movie always arrives in the count's pace.

4. in dark clouds (in huge/excessive amount)
 - (a) **Control:** During the holiday season, everyone forgets their diets and eats in dark clouds.
 - (b) **Adjective Insertion:** During the holiday season, everyone forgets their diets and eats in dark sweet clouds.
 - (c) **Scalar Substitution:** During the holiday season, everyone completely ignores their diets and eats in black clouds.
 - (d) **Non-scalar Substitution:** During the holiday season, everyone forgets their diets and eats in shadowy clouds.
5. on the hot porridge (very eager)
 - (a) **Control:** The college freshman was so excited that she signed up for ten different extracurricular activities; when it came to college, she was on the hot porridge.
 - (b) **Adjective Insertion:** The college freshman was so excited that she signed up for ten different extracurricular activities; when it came to college, she was on the hot social porridge.
 - (c) **Scalar Substitution:** The college freshman was ambivalent and didn't sign up for any extracurricular activities; her RA was concerned that she was only on the cool porridge.
 - (d) **Non-scalar Substitution:** The college freshman was so excited that she signed up for ten different extracurricular activities; when it came to college, she was on the hot oatmeal.
6. on the wagon (intelligent, put together)
 - (a) **Control:** Ms. Jones was not only a gifted teacher, but she was extremely knowledgeable about math; when it came to teaching, she was really on the wagon.
 - (b) **Adjective Insertion:** Ms. Jones was not only a gifted teacher, but she was extremely knowledgeable about math; when it came to teaching, she was really on the mathematical wagon.
 - (c) **Scalar Substitution:** Ms. Jones was neither good at teaching nor knowledgeable about math; when it came to being a math teacher, she was really off the wagon.
 - (d) **Non-scalar Substitution:** Ms. Jones was not only a gifted teacher, but she was extremely knowledgeable about math; when it came to teaching, she was really on the carriage.
7. against the walls (completely wrong)
 - (a) **Control:** The TV Show set in the 1800s uses lots of phrases that weren't around then; when it comes to time-period accuracy, the screenwriters were against the walls.
 - (b) **Adjective Insertion:** The TV Show set in the 1800s uses lots of phrases that weren't around then; when it comes to time-period accuracy, the screenwriters were against the linguistic walls.
 - (c) **Scalar Substitution:** The TV Show set in the 1800s, while mostly accurate, uses occasional phrases that weren't around then; when it comes to time-period accuracy, the screenwriters were near the walls.

- (d) **Non-scalar Substitution:** The TV Show set in the 1800s uses lots of phrases that weren't around then; when it comes to time-period accuracy, the screenwriters were touching the walls.
8. with helmet askew (downhearted)
- (a) **Control:** After the difficult breakup, Tim just lay around his room with helmet askew.
- (b) **Adjective Insertion:** After the difficult breakup, Tim just lay around his room with relational helmet askew.
- (c) **Scalar Substitution:** After the difficult breakup, Tim's roommate was surprised to find him lying around his room with helmet on straight.
- (d) **Non-scalar Substitution:** After the difficult breakup, Tim just lay around his room with helmet awry.
9. with one's beard in the mailbox ((caught) by surprise)
- (a) **Control:** Derrin was cleaning up his trashed house after the party when his grandmother came to visit; she found him with his beard in the mailbox.
- (b) **Adjective Insertion:** Derrin was cleaning up his trashed house after the party when his grandmother came to visit; she found him with his beard in the domestic mailbox.
- (c) **Scalar Substitution:** Derrin was just about finished cleaning up his trashed house after the party when his grandmother came to visit; she found him with just his mustache in the mailbox.
- (d) **Non-scalar Substitution:** Derrin was cleaning up his trashed house after the party when his grandmother came to visit; she found him with his facial hair in the mailbox.
10. in a curious cone (very nosy)
- (a) **Control:** Because he kept asking Lily and Sam about the details of their relationship, the couple accused John of being in a curious cone.
- (b) **Adjective Insertion:** Because he kept asking Lily and Sam about the details of their relationship, the couple accused John of being in a curious gossipy cone.
- (c) **Scalar Substitution:** When he didn't even ask about his best friends' engagement, the couple accused John of being in an indifferent cone.
- (d) **Non-scalar Substitution:** Because he kept asking Lily and Sam about the details of their relationship, the couple accused John of being in a prying cone.

A.2.3 NP

1. ice in the stomach (control of a situation)
- (a) **Control:** Although the class was unruly with substitutes, the teacher was able to keep them under control, since he had ice in the stomach.
- (b) **Adjective Insertion:** Although the class was unruly with substitutes, the teacher was able to keep them under control, since he had ice in the teaching stomach.

- (c) **Scalar Substitution:** Although the class was unruly with substitutes, the teacher was able to keep them under control, since he had ice in his whole torso.
 - (d) **Non-scalar Substitution:** Although the class was unruly with substitutes, the teacher was able to keep them under control, since he had ice in the belly.
2. fly leg in the universe (something inconsequential)
- (a) **Control:** Considering all of the terrible plays in the football game, the incomplete pass was just a fly leg in the universe.
 - (b) **Adjective Insertion:** Considering all of the terrible plays in the football game, the incomplete pass was just a fly leg in the playmaking universe.
 - (c) **Scalar Substitution:** Considering all of the terrible plays in the football game, the incomplete pass was just a tiny fly foot in the universe.
 - (d) **Non-scalar Substitution:** Considering all of the terrible plays in the football game, the incomplete pass was just a fly leg in the cosmos.
3. someone's heavy coin (slip of the tongue)
- (a) **Control:** When Anna ordered a small pizza instead of a large one, she immediately corrected her mistake; it was just her heavy coin.
 - (b) **Adjective Insertion:** When Anna ordered a small pizza instead of a large one, she immediately corrected her mistake; it was just her heavy thick coin.
 - (c) **Scalar Substitution:** When Anna ordered a small pizza instead of a large one, she immediately corrected her tiny mistake, since it was just her light coin.
 - (d) **Non-scalar Substitution:** When Anna ordered a small pizza instead of a large one, she immediately corrected her mistake; it was just her weighty coin.
4. cake on top of cake (redundant, too much of a good thing)
- (a) **Control:** Ashley found that vacationing in Jamaica and then the Bahamas was really cake on top of cake.
 - (b) **Adjective Insertion:** Ashley found that vacationing in Jamaica and then the Bahamas was really cake on top of delicious cake.
 - (c) **Scalar Substitution:** Ashley found that vacationing at the Jersey shore after her trip to the Bahamas was really a cupcake on top of cake.
 - (d) **Non-scalar Substitution:** Ashley found that vacationing in Jamaica and then the Bahamas was really cake added to cake.
5. a bridge of a donkey (tenuous connection between discussion topics)
- (a) **Control:** The news anchor's transition from the fiscal cliff to Kim Kardashian's pregnancy was bound to be a bridge of a donkey.
 - (b) **Adjective Insertion:** The news anchor's transition from the fiscal cliff to Kim Kardashian's pregnancy was bound to be a bridge of a feeble donkey.
 - (c) **Scalar Substitution:** The news anchor managed to make his transition from the fiscal cliff to Kim Kardashian's pregnancy to be a bridge of a horse.

- (d) **Non-scalar Substitution:** The news anchor’s transition from the fiscal cliff to Kim Kardashian’s pregnancy was bound to be a bridge of a mule.
6. a chicken to pluck ([to have] unresolved disputes)
- (a) **Control:** After Patrick crashed his friend’s car, the two of them had a chicken to pluck.
- (b) **Adjective Insertion:** After Patrick crashed his friend’s car, the two of them had a monetary chicken to pluck.
- (c) **Scalar Substitution:** After Patrick crashed his friend’s brand new luxury car, the two of them had a huge ostrich to pluck.
- (d) **Non-scalar Substitution:** After Patrick crashed his friend’s car, the two of them had a hen to pluck.
7. a dog buried somewhere (something foul behind the scenes)
- (a) **Control:** Looking at the company’s sketchy records and meeting with the evasive executives, Angela couldn’t help but feel that there was a dog buried there.
- (b) **Adjective Insertion:** Looking at the company’s sketchy records and meeting with the evasive executives, Angela couldn’t help but feel that there was a decaying dog buried there.
- (c) **Scalar Substitution:** Looking at the company’s sketchy records and meeting with the evasive executives, Angela couldn’t help but feel that there was a huge wolf buried there.
- (d) **Non-scalar Substitution:** Looking at the company’s sketchy records and meeting with the evasive executives, Angela couldn’t help but feel that there was a hound buried there.
8. snows of past winters (something no longer relevant)
- (a) **Control:** When they met up twenty years later, the athletes’ high school basketball rivalry hardly mattered; it was just snows of past winters.
- (b) **Adjective Insertion:** When they met up twenty years later, the athletes’ high school basketball rivalry hardly mattered; it was just snows of past competitive winters.
- (c) **Scalar Substitution:** When they met up twenty years later, the athletes’ minor high school basketball rivalry hardly mattered; it was just flurries of past winters.
- (d) **Non-scalar Substitution:** When they met up twenty years later, the athletes’ high school basketball rivalry hardly mattered; it was just snows of previous winters.
9. brother-in-law politics (favoritism)
- (a) **Control:** Most of the people on the president’s staff were his friends from college, so everyone suspected that they had gotten their jobs through brother-in-law politics.
- (b) **Adjective Insertion:** Most of the people on the president’s staff were his friends from college, so everyone suspected that they had gotten their jobs through brother-in-law hiring politics.

- (c) **Scalar Substitution:** Most of the people on the president's staff were his friends from college or friends of his parents, so everyone suspected that they had gotten their jobs through family-in-law politics.
 - (d) **Non-scalar Substitution:** Most of the people on the president's staff were his friends from college, so everyone suspected that they had gotten their jobs through sister-in-law politics.
10. nose-bleaching (sobering up)
- (a) **Control:** After Manny took eight shots on St. Patrick's day, he needed a serious nose-bleaching.
 - (b) **Adjective Insertion:** After Manny took eight shots on St. Patrick's day, he needed a serious drunken nose bleaching.
 - (c) **Scalar Substitution:** After Manny took eight shots on St. Patrick's day, he was so drunk that he needed an entire face-purging.
 - (d) **Non-scalar Substitution:** After Manny took eight shots on St. Patrick's day, he needed a serious nose-purging.

Appendix B

Results of Experiment 1

This appendix contains the results from Experiment 1. The first table contains the results for real idioms, and the second table contains the results for novel idioms. For each idiom, and each type of modification, both the mean judgment value and the standard deviation are provided. Below is a key to the abbreviations for each condition:

- **Con:** Control (“break the ice”)
- **Adj:** Adjective insertion (“break the stuffy ice”)
- **Syn:** Non-scalar substitution (“split the ice”)
- **ScI:** Scalar substitution (“shatter the ice”)
- **Psv:** Passivization (“the ice was broken”)

In the column labels, these condition abbreviations are combined with either “M” for “mean” or “SD” for standard deviation.

Idiom	ConM	AdjM	SynM	ScIM	PsvM	ConSD	AdjSD	SynSD	ScISD	PsvSD
slip someone's mind	6.97	5.58	3.86	3.42	1.5	0.19	1.91	1.84	2.08	1.1
hit the sack	6.92	4.67	2.24	1.68	2.32	0.28	1.83	1.83	0.99	1.49
play with fire	6.54	4.38	3.73	4.42	2.54	1.02	2.08	1.8	1.77	1.69
cost an arm and a leg	7	5.04	6.28	5.91	2.79	0	1.99	1.1	1.85	1.72
swallow one's pride	6.96	5.54	3.67	2.72	3.59	0.2	1.69	2.04	1.67	1.68
face the music	6.63	5.03	3.82	2.75	3.61	0.97	2.04	1.79	1.67	1.87
get the picture	6.82	5.46	3.21	3.08	3.72	0.39	1.61	1.64	2.04	2.22
jump the gun	6.91	5.75	2.71	3.54	3.76	0.43	0.99	1.65	1.91	2.12
tie the knot	7	4.97	3.27	4.54	5.13	0	2.2	1.8	2.08	1.83
break the ice	6.86	3.59	3.67	4.46	5.21	0.74	1.59	1.81	1.84	1.44
twist someone's arm	6.36	4.71	3.25	2.24	5.67	1.65	2.12	2.13	1.6	1.69
call the shots	7	6.25	3.14	3	5.75	0	1.03	2.08	1.9	1.89
learn by heart	6.83	3.73	6.29	2.32	6.25	0.93	1.98	0.95	1.52	1.03
give someone the cold shoulder	6.92	3.5	4.17	4.07	6.83	0.28	1.74	1.66	1.72	0.54
a frog in one's throat	6.62	4.64	3.71	2.46	NA	1.18	2.01	1.97	1.82	NA
a horse of another color	6	4.41	4.46	2.75	NA	1.53	1.76	1.91	1.96	NA
a piece of cake	6.75	4.93	3.77	2.96	NA	1.03	1.98	2.05	1.78	NA
at the back of one's mind	6.43	5.04	4	1.64	NA	1.17	1.4	1.98	1.14	NA
fish out of water	6.95	4.04	4.38	3.66	NA	0.21	1.88	1.81	2.22	NA
food for thought	6.86	5.67	3.86	3.54	NA	0.64	1.49	1.99	2.11	NA
in hot water	6.79	3.91	4.54	4.39	NA	0.59	1.44	1.89	1.87	NA
in seventh heaven	6.91	4.54	2.93	3.29	NA	0.29	1.86	1.84	2.33	NA
in the nick of time	6.96	4.11	3.17	1.9	NA	0.2	1.69	2.18	1.52	NA
on thin ice	6.82	4.14	3.63	2.48	NA	0.85	2.17	2.12	1.96	NA
out of the blue	6.79	3.71	2.07	1.96	NA	0.82	2.1	1.51	1.6	NA
out of thin air	6.08	4.13	1.93	2.55	NA	1.35	2.09	1.31	1.74	NA
over the hill	6.17	5	3.5	2.68	NA	1.61	1.8	2.02	1.72	NA
par for the course	6.67	4.39	3.04	1.79	NA	1.05	1.95	2.22	1.5	NA
rule of thumb	7	4.76	3.18	3.58	NA	0	2.18	1.74	1.91	NA
tip of the iceberg	6.92	5.38	3.86	3.36	NA	0.28	1.86	2.22	2.26	NA
two peas in a pod	6.52	5.83	3.04	3.38	NA	0.99	1.24	1.93	2.2	NA
under the weather	6.96	3.07	2.77	1.79	NA	0.2	2.02	1.8	1.32	NA
with flying colors	7	4.9	2	4.5	NA	0	2.27	1.11	2.3	NA

Idiom	ConM	AdjM	SynM	ScIM	PsvM	ConSD	AdjSD	SynSD	ScISD	PsvSD
speak out of the beard	5.03	2.63	4.64	3.29	1.45	2.47	1.76	2.20	1.90	1.18
twist from wire	5.42	3.42	1.90	1.91	1.46	1.95	2.26	1.54	1.27	0.88
hit a drill	5.41	4.58	2.96	3.08	1.66	2.42	2.02	1.97	2.21	1.17
begin on a new marble	5.79	4.54	4.38	2.28	1.95	2.06	1.86	2.32	1.91	1.46
walk on the pump	5.43	2.83	2.17	2.32	1.96	2.06	1.95	1.54	1.73	1.46
roll hat	4.88	3.03	1.73	2.29	2.32	2.05	1.72	1.32	1.46	1.70
play one's mouth	5.61	3.13	1.83	1.41	2.33	1.87	1.65	1.58	0.85	1.37
walk in the roof	5.58	1.82	2.75	2.04	2.45	2.00	1.44	1.87	1.50	2.01
drop the suck	5.59	4.75	1.88	2.72	2.58	2.40	2.08	1.39	1.94	1.64
grasp the gallop	5.83	4.50	3.75	2.59	2.77	1.88	2.01	2.19	1.94	1.69
put onion on the salmon	6.05	4.25	5.36	1.96	2.83	1.99	2.01	1.91	1.85	2.17
throw a goat's eye	5.00	4.72	3.59	1.42	2.86	2.02	2.23	2.24	1.28	2.16
throw the spoon	5.83	2.82	4.42	3.93	3.33	1.69	1.71	2.00	2.18	2.12
put on someone's fallen cloak	6.00	3.68	5.50	2.54	4.71	1.49	1.94	1.44	1.67	2.24
a bridge of a donkey	5.18	3.88	2.50	2.04	NA	2.40	2.17	1.86	1.78	NA
a chicken to pluck	5.75	4.50	4.25	2.57	NA	1.51	2.30	1.65	1.69	NA
a dog buried somewhere	5.21	3.83	4.14	2.64	NA	2.17	2.43	2.20	1.71	NA
against the walls	5.17	4.54	2.08	2.07	NA	2.06	2.24	1.64	1.73	NA
behind the floating	4.72	3.32	1.92	2.29	NA	2.48	2.10	1.21	1.94	NA
brother-in-law politics	6.21	5.79	2.66	3.95	NA	1.44	1.10	1.88	2.15	NA
cake on top of cake	5.90	4.71	4.71	3.58	NA	1.92	2.24	2.11	2.28	NA
fly leg in the universe	5.67	4.89	3.88	2.83	NA	2.04	2.13	2.15	2.14	NA
ice in the stomach	4.95	3.71	3.79	2.07	NA	2.36	2.07	2.04	1.89	NA
in a curious cone	5.54	3.79	2.21	1.45	NA	2.11	2.19	1.57	1.01	NA
in dark clouds	5.11	2.90	1.86	3.17	NA	2.22	2.30	1.55	1.86	NA
in the count's time	5.76	4.96	3.32	1.38	NA	2.05	1.90	2.02	1.17	NA
nose-bleaching	5.76	3.45	1.92	1.75	NA	2.08	1.71	1.35	1.29	NA
on the hot porridge	5.71	3.38	2.00	2.21	NA	1.52	1.86	1.60	1.50	NA
on the wagon	5.95	5.21	2.92	3.03	NA	2.03	1.97	2.02	2.26	NA
on the wallpaper	5.93	3.88	1.03	3.18	NA	1.98	2.44	0.19	2.17	NA
snows of past winters	6.17	4.28	5.27	4.25	NA	1.46	2.31	1.72	1.96	NA
someone's heavy coin	4.00	2.34	2.45	1.54	NA	2.34	1.88	1.77	1.22	NA
with helmet askew	4.91	3.92	3.43	1.92	NA	2.41	2.30	2.06	1.50	NA
with one's beard in the mailbox	4.63	2.59	2.50	2.36	NA	2.10	1.56	1.69	1.85	NA

Appendix C

Plausibility for Novel Idioms in Experiment 1

C.1 Overview

This appendix explains the methodology and provides the results of a study to determine how plausible the novel idioms used in Experiment 1 were. This was used to analyze the flexibility results in Experiment 1 (see Chapter 2).

C.1.1 Participants

Participants were recruited through Amazon Mechanical Turk and were paid \$0.25 for an average of 8 minutes and 48 seconds of work. Participants had a HIT approval rate of 95% or above, had completed at least 1000 HITs, and had IP addresses from North America. In the week the survey was available, 45 participants responded.

C.1.2 Materials and Procedure

All of the novel idioms used in Experiment 1 were presented, along with all of the real idioms used in Experiment 1. Participants' task was to determine the plausibility of the idiom on a scale from 1 (less plausible) to 7 (more plausible). They were instructed to select "Real idiom" instead of a number if the idiom was real (the filler quality-control task). Plausibility was explained using the following directions:

Select "1" if you think it would be impossible for such an idiom to exist in English. Select "7" if you would be able to learn and begin to use it without difficulty, and select an appropriate number between 1 and 7 to reflect intermediate levels of plausibility (closer to 1= less plausible; closer to 7 = more plausible).

Each participant rated every idiom, and idioms were presented in a random order.

C.2 Results

C.2.1 Quality Control

Participants' data were excluded if any of the following conditions were met: survey completion time or instructional reading time was greater or less than 1.5 standard deviations from the mean, participants identified novel idioms as real more than once. After these criteria were applied, 25 participants remained.

C.2.2 Plausibility Scores

Mean and medians scores on the 1-7 scale were calculated for each idiom. These were used in the analysis in Experiment 1 and can be seen in Table C.1.

Table C.1: Plausibility Scores for the Novel Idioms in Experiment 1

Idiom	Mean	Median
a bridge of a donkey	2.38	2
a chicken to pluck	4.53	5
a dog buried somewhere	3.86	5
against the walls	3.24	2
behind the floating	1.76	1
brother-in-law politics	4.68	5
cake on top of cake	4.58	5
fly leg in the universe	2.33	2
ice in the stomach	2.52	2
in a curious cone	2.81	2
in dark clouds	2.57	2
in the count's time	3.05	2
nose-bleaching	2.00	2
on the hot porridge	2.67	2
on the wagon	3.50	3
on the wallpaper	2.48	2
snows of past winters	5.26	6
someone's heavy coin	2.05	1
to begin on a new marble	2.62	2
to drop the suck	1.90	1
to give someone the rake	2.57	3
to grasp the gallop	3.05	3
to hit a drill	2.19	1
to play one's mouth	3.19	3
to put on someone's fallen cloak	3.86	4
to put onion on the salmon	2.90	2
to roll hat	2.38	2
to speak out of the beard	2.67	2
to throw a goat's eye	2.86	2
to throw the spoon	2.52	2
to twist from wire	2.90	3
to walk in the roof	1.76	1
to walk on the pump	2.43	2
with helmet askew	3.38	3
with one's beard in the mailbox	2.90	2

Bibliography

- Behrend, Tara S, David J Sharek, Adam W Meade, and Eric N Wiebe. 2011. The viability of crowdsourcing for survey research. *Behavior research methods* 43:800–13. URL <http://www.ncbi.nlm.nih.gov/pubmed/21437749>.
- Butt, Miriam. 2003. The Light Verb Jungle. *Workshop on Multi-Verb Constructions* 1–28. URL <http://ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/papers/harvard-work.pdf>.
- Cacciari, Cristina, and Patrizia Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language* 27:668–683. URL <http://linkinghub.elsevier.com/retrieve/pii/S0749596X88900149>.
- Caillies, Stéphanie, and Kirsten Butcher. 2007. Processing of Idiomatic Expressions: Evidence for a New Hybrid View. *Metaphor and Symbol* 22:79–108. URL http://www.leaonline.com/doi/abs/10.1207/s15327868ms2201_3.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. New York: Foris.
- Cutting, J C, and K Bock. 1997. That’s the way the cookie bounces: syntactic and semantic components of experimentally elicited idiom blends. *Memory & cognition* 25:57–71. URL <http://www.ncbi.nlm.nih.gov/pubmed/9046870>.
- Espinal, M. Teresa, and Jaume Mateu. 2010. On classes of idioms and their interpretation. *Journal of Pragmatics* 42:1397–1411. URL <http://linkinghub.elsevier.com/retrieve/pii/S0378216609002471>.
- Fraser, Bruce. 1970. Idioms within a Transformational Grammar. *Foundations of language* 6:22–42. URL <http://www.jstor.org/stable/25000426>.
- Gibbs, R W. 1980. Spilling the beans on understanding and memory for idioms in conversation. *Memory & cognition* 8:149–56. URL <http://www.ncbi.nlm.nih.gov/pubmed/7382816>.
- Gibbs, Raymond W. 1995. Idiomaticity and Human Cognition. *Idioms: Structural and psychological perspectives* 97–116. URL <http://books.google.com/books?hl=en&lr=&id=mbAKf2b1K9MC&oi=fnd&pg=PA97&dq=Idiomaticity+and+Human+Cognition&ots=9kTAfad7x0&sig=o3FDUVEUmcaGFkKr7UuyCryc9ls>.
- Gibbs, Raymond W., and Nadini P. Nayak. 1989. Psycholinguistic Studies on the Syntactic Behavior of Idioms. *Cognitive Psychology* 21:100–138. URL <http://linkinghub.elsevier.com/retrieve/pii/0010028589900042>.

- Gibbs, Raymond W., Nadini P. Nayak, J L Bolton, and M E Keppel. 1989. Speakers' assumptions about the lexical flexibility of idioms. *Memory & cognition* 17:58–68. URL <http://www.ncbi.nlm.nih.gov/pubmed/17363195>.
- Gibbs, RW, and JE O'Brien. 1990. Idioms and Mental Imagery: The Metaphorical Motivation for Idiomatic Meaning. *Cognition* 35–68. URL <http://www.sciencedirect.com/science/article/pii/001002779090053M>.
- Glucksberg, Sam. 2001. *Understanding Figurative Language: From Metaphor to Idioms*. Cambridge: Cambridge University Press.
- Jackendoff, Ray. 1995. The Boundaries of the Lexicon. In *Idioms: Structural and psychological perspectives*, ed. Martin Everaert, Erik-Jan van der Linden, André Schenk, and Rob Schreuder, chapter 7, 133–165. Hillsdale: Lawrence Erlbaum Associates.
- Jackendoff, Ray. 1997. Idioms and Other Fixed Expressions. In *The architecture of the language faculty*, volume 28, chapter 7. MIT Press. URL http://books.google.com/books?hl=en&lr=&id=_AkHjNi6qqcC&oi=fnd&pg=PR11&dq=The+architecture+of+the+language+faculty&ots=6gOfRfmsyj&sig=w96zQSfx7ikfeHXtNYblxL_NeuI.
- Jackendoff, Ray. 2002. Lexical Storage versus Online Construction. In *Foundations of language*, January, chapter 6, 152–195. Oxford University Press. URL <http://psycnet.apa.org/psycinfo/1940-00426-000>.
- Jespersen, Otto. 1965. *A Modern English Grammar on Historical Perspectives, Part VI, Morphology*. London: George Allen and Unwin Ltd.
- Kearns, K. 2002. Light verbs in English 1–33. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.29&rep=rep1&type=pdf>.
- Lakoff, George, and Mark Johnson. 1980. The Metaphorical Structure of the Human Conceptual System. *Cognitive Science* 4:195–208. URL http://doi.wiley.com/10.1207/s15516709cog0402_4.
- Marslen-Wilson, W D. 1987. Functional parallelism in spoken word-recognition. *Cognition* 25:71–102. URL <http://www.ncbi.nlm.nih.gov/pubmed/18480284>.
- Mason, Winter, and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44:1–23. URL <http://www.ncbi.nlm.nih.gov/pubmed/21717266>.
- McGlone, Matthew, Sam Glucksberg, and Cristina Cacciari. 1994. Semantic productivity and idiom comprehension. *Discourse Processes* 17:167–190. URL <http://www.informaworld.com/openurl?genre=article&doi=10.1080/01638539409544865&magic=crossref||D404A21C5BB053405B1A640AFFD44AE3>.
- Munro, Robert, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai, Robin Melnick, Christopher Potts, Tyler Schnoebelen, and Harry Tily. 2010. Crowdsourcing and language studies: the new generation of linguistic data. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, 122–130. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=1866715>.

- North, Ryan. 2005. Computational Measures of the Acceptability of Light Verb Constructions. Doctoral Dissertation, University of Toronto.
- Nunberg, Geoffrey. 1978. *The pragmatics of reference*. Indiana University Linguistics Club. URL <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+pragmatics+of+reference#0>.
- Nunberg, Geoffrey, Ivan A. Sag, and Thomas Wasow. 1994. Idioms. *Language* 70:491–538. URL <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Idioms#1>.
- Paolacci, Gabriele, Jesse Chandler, and P Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5:411–419. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1626226.
- R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <http://www.r-project.org>.
- Rabanus, Stefan, Eva Smolka, Judith Streb, and Frank Rösler. 2008. Die mentale Verarbeitung von Verben in idiomatischen Konstruktionen. *Zeitschrift für Germanistische Linguistik* 36:27–47. URL <http://www.reference-global.com/doi/abs/10.1515/ZGL.2008.003>.
- Smolka, Eva, Stefan Rabanus, and Frank Rösler. 2007. Processing verbs in German idioms: Evidence against the Configuration Hypothesis. *Metaphor and Symbol* 22:213–231. URL <http://www.informaworld.com/index/787952529.pdf>.
- Sprenger, Simone A., Willem J.M. Levelt, and Gerard Kempen. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54:161–184. URL <http://linkinghub.elsevier.com/retrieve/pii/S0749596X05001300>.
- Swinney, David A. 1979. Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18:645–659. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022537179903554>.
- Tabossi, Patrizia, Rachele Fanari, and K. Wolf. 2009a. Why are idioms recognized fast? *Memory & cognition* 37:529–540. URL <http://www.springerlink.com/index/93U223415032L058.pdf>.
- Tabossi, Patrizia, Rachele Fanari, and Kinou Wolf. 2008. Processing idiomatic expressions: effects of semantic compositionality. *Journal of experimental psychology. Learning, memory, and cognition* 34:313–27. URL <http://www.ncbi.nlm.nih.gov/pubmed/18315408>.
- Tabossi, Patrizia, Kinou Wolf, and Sara Koterle. 2009b. Idiom syntax: Idiosyncratic or principled? *Journal of Memory and Language* 61:77–96. URL <http://linkinghub.elsevier.com/retrieve/pii/S0749596X09000369>.
- Titone, Debra A., and Cynthia M Connine. 1994. Descriptive Norms for 171 Idiomatic Expressions. *Metaphor and Symbol* 9:247–270. URL <http://www.informaworld.com/index/785872582.pdf>.
- Titone, Debra A., and Cynthia M Connine. 1999. On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics* 31:1655–1674.